(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
13 June 2002 (13.06.2002)

PCT

(10) International Publication Number
WO 02/46439 A2

(51) International Patent Classification⁷: C12N 15/82, 15/29, 5/10, 9/12, A01H 5/00, G01N 33/53

(21) International Application Number: PCT/EP01/14154

(22) International Filing Date: 4 December 2001 (04.12.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
00870288.8 4 December 2000 (04.12.2000) EP
60/256,204 15 December 2000 (15.12.2000) US

(71) Applicant (*for all designated States except US*): UNIVERSITEIT UTRECHT [NL/NL]; Heidelberglaan 8, NL-3584 CS Utrecht (NL).

(72) Inventors; and

(75) Inventors/Applicants (*for US only*): SCHERES, Ben [NL/NL]; Vlieland 106, NL-3524 AC Utrecht (NL). HEIDSTRA, Renze [NL/NL]; Hellendoornseberg 13, NL-3524 LH Utrecht (NL).

(74) Agents: DE CLERCQ, Ann et al.; De Clercq, Brants & Partners, E. Gevaertdreef 10a, B-9830 Sint-Martens-Latem (BE).

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— *without international search report and to be republished upon receipt of that report*

— *entirely in electronic form (except for this front page) and available upon request from the International Bureau*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 02/46439 A2

(54) Title: A NOVEL ROOT SPECIFIC PROMOTER DRIVING THE EXPRESSION OF A NOVEL LRR RECEPTOR-LIKE KINASE

(57) Abstract: The present invention relates to the field of plant molecular biology, more particularly to the root-specific gene expression in plants. The invention provides nucleic acids for a novel transcriptional regulatory root-specific promoter and nucleic acid and protein sequences coding for a new LRR receptor-kinase protein, further specified as a root clavata 1 homolog (RCH1). Further provided are compositions comprising nucleic acids, polypeptides, antibodies and vectors. The invention further provides for methods for modifying cell fate and/or plant development and/or plant morphology and/or plant biochemistry and/or plant physiology comprising the modification of expression in particular cells, tissues or organs of a plant of the novel LRR receptor-like kinase or comprising the expressing of a gene of interest under the control of the novel transcriptional regulatory root-specific promoter. Further are provided compounds interacting with the new polypeptides for use as herbicides or growth regulators.

A NOVEL ROOT SPECIFIC PROMOTER DRIVING THE EXPRESSION OF A NOVEL LRR RECEPTOR-LIKE KINASE

FIELD OF THE INVENTION

The present invention relates to the field of plant molecular biology, more particularly to the root-specific gene expression in plants. The isolation of a root specific operon, comprising a transcriptional regulatory promoter that contributes to root-specific gene expression and its operably linked gene encoding a novel Leucin Rich Repeat (LRR) receptor-like kinase, is disclosed. Said transcriptional regulatory promoter may be used for driving root-specific expression of at least one gene of interest in a transgenic plant and said encoded LRR receptor-like kinase gene may be used to alter the features and/or to confer a selective advantage to transgenic plants.

BACKGROUND TO THE INVENTION

Plant promoters

Initiation of transcription is generally understood to be the predominant controlling factor in determining expression of a gene. The transcriptional control elements, which may interact with DNA binding proteins, are generally embedded in the sequence 5'-flanking or upstream of the transcribed gene. These DNA sequence elements promote the formation of transcriptional regulatory complexes that either activate or repress the expression of the gene 3' downstream of the promoter.

The sequences of the promoter vary in length and base pair composition from gene to gene. Within the sequence of the promoter, many regulatory sequence elements can be embedded that control the promoter activity. These specific elements, called DNA-boxes, contribute to the promoter a defined feature or a defined activation pattern. Many of these boxes are recognition sites to which regulatory transcription factors can bind and are part of a tight control mechanism of the promoter. Numerous plant promoter DNA boxes have been described, such as tissue-specific boxes (Chaubet et al. 1996), pyrimidin boxes (Gubler & Jacobsen 1992), which influence the level of expression or G-boxes (Dolferus et al. 1994) which reduce promoter activity by cold or dehydratation.

Constitutive promoters are promoters that drive the expression of genes in a continuous manner, whereas ubiquitous promoters drive the expression of the gene throughout the entire plant.

Expression of heterologous DNA sequences in a host is dependent upon the presence of an operably linked promoter that is also present in that host, whereby the choice of promoter will determine when, where or how strong the heterologous DNA sequence will be expressed in the host organism. Frequently those skilled in the art desire to limit the expression of the transfected DNA sequence to a certain time-period (e.g. to a particular phase of plant development), to a defined tissue of the host (e.g. leaves), or to a certain expression level. Controlling the expression of the heterologous genes in transgenic plants is considered to provide several advantages to the plant over ubiquitous and constitutive expression. There is a broad need in many industrial applications for transcriptional control elements capable of driving tissue specific gene-expression in plants. It is therefore important that such control elements are continuously provided and that new control elements with their specific transcriptional features are isolated and characterized.

Promoter regulation

Promoters from both monocots and dicots that have a root root-preferred expression pattern have already been described (Fowler, Bernhardt, & Tierney 1999;Hwang & Goodman 1995;Lagarde et al. 1996;Okada et al. 2000;Okumura et al. 1994;Pysh et al. 1999.Schneider et al. 1998;Vidmar et al. 2000;Wanapu & Shinmyo 1996;Winicov 2000)WO9007001(Baszczynski et al. 2000;Borg et al. 1997;Masuda, Sakuta, & Satoh 1999.Webb et al. 2000) (de Pater & Schilperoort 1992;Ely, Evans, & Schuch 2000.Goddemeier, Wulff, & Feix 1998;Kikuchi et al. 1999;Suzuki, Fowler, & Tierney 1993.Unbe et al. 1998;Woo & Hawes 1997;Xu et al. 1995) (Yamamoto et al. 1991) (Gritch et al. 1997;Keller & Lamb 1989;Lauter 1996;Olson, Oetiker, & Yang 1995;van der Zaal et al. 1991). Many of them are root-preferred rather then root-specific and are expressed at low levels in tissues other than the root. These promoters are not suitable for industrial applications where it us of the utmost importance that the heterologous product is expressed in the root only. As an example, ectopic expression of cell cycle control elements to obtain growth arrest in some parts of the root (e.g. lateral roots),

should be strictly regulated to that tissue in order not to interfere with the overall growth of the plant. In contrast, other promoters are strictly and exclusively confined to root-specific expression and can therefore be valuable for some applications demanding tight control of the expression. In this group of root-specific promoter subgroups can be distinguished based on the root-tissue they are expressed in. The HRGPnt3 gene of tobacco for instance is only expressed in the pericycle and the endodermis of the root, more specifically in the discrete subset of cells, involved in the initiation of the lateral roots (Keller et al. 1989). This makes the promoter of HRGPnt3 highly specific. Another example is the XSP30 sequence, which is only activated in the xylem of the root.

For persons skilled in the art it is of importance to have at their disposal a wide variety of root-specific promoters with different root tissue-specificity from which they can choose to confer a specific feature to the transgenic plant.

Many of these root-preferred promoters are not only limited to tissue specific expression, but are also subjected to a tight control mechanism deciding when the promoter should be active. This decision depends on the function and the effects of the gene product that is switched on/off by the related promoter. Some gene products are switched on/off at a certain developmental stage of the plant, at a certain moment of the cell cycle, in stress situations originating from weather factors, upon injury caused by pathogens or upon a change in the environmental resources. A typical example is the ARSK1 gene of *Arabidopsis thaliana*, that is activated when the roots are exposed to air during growth (dehydration) or by treatment of roots with abcisic acid or salt (Hwang et al. 1995).

However, for many industrial and agronomic applications, where the aim is to confer a special feature at any time of the developmental, environmental or physical state of the plant, a constitutive promoter will be needed. For example when the gene product of the transgene is to be isolated and purified from the plant for commercial purposes.

It is important for the persons skilled in the art to have at their disposal a wide variety of promoters with different time-control features from which they can chose to confer a specific feature to the transgenic plant.

Plant root development

Root development is an essential determinant of plant growth and crop yield since the root is the main channel to extract nutrients from the environment. Any changes in the

root that contribute to more functional organ tissue will result in higher crop yield. Also any regulatory element that is involved in the root development and whose control is subjected to developmental events can contribute better yield properties to a transgenic plant.

Cells of the shoot and root meristems are mitotically active and they extend the undifferentiated tissue pattern of the embryo as a pool of cells from which cells can differentiate into several tissues of the adult plant. This process continues throughout the life span of the plant. It is clear that the root meristem plays a pivotal role in the formation of strong roots and therefor also in affecting the growth and crop yield.

In root meristem, the emerging picture is that the integrity of the root meristem is kept by balancing cell proliferation and cell differentiation, and that differentiation-inhibiting signals originating from the quiescent center are involved (van den Berg et al. 1997). Inhibition of differentiation only occurs in cells that contact the quiescent center, which reveals that short-range or contact-dependent signals are involved in this process.

However which compounds play a role in this process and whether positional cues guiding root cell fate are targeted via plasmodesmata or via the cell wall remains unknown (van den Berg et al. 1997).

In the shoot meristem more progress has been made in characterizing molecular components involved in meristem maintenance. *Clavata (clv)* mutants accumulate massive pools of undifferentiated cells in the central zone of shoot and floral meristem. The *CLAVATA* genes show a shoot and floral meristem specific expressionpattern. Root meristems of *clv*-mutants are unaffected (Clark, Running, & Meyerowitz 1993). The *CLAVATA* loci (*CLV1*, *CLV2* and *CLV3*) appear to promote the transition toward differentiation of cells at the shoot and floral meristems, and/or to restrict the proliferation of cells at the center of these meristems (Clark et al. 1993). *CLV1* acts with *CLV3* and WUSCHEL (WUS) and *SHOOT MERISTEMLESS (STM)* to maintain the integrity of the central zone which acts as a reservoir of stem cells. A model for stem-cell maintenance in the root meristem and in the shoot meristem is depicted in Figure 1.

CLV1 homologs and CLV1-redundancy

The *CLAVATA1* gene resides on the *Arabidopsis thaliana* chomosome 1. *Clv1* mRNA is found in the L3 tunica cell layer of the central zone and encodes a protein with leucine-rich repeats (LRR, of a type found in a number of proteins, notably the products of

several plant-pathogen resistance genes), a putative serin/threonin kinase domain and a predicted extracellular receptor domain. Expression of the *clv1* gene is initiated in the heart stage of the embryo, when cotyledonary primordia are apparent, and is independent of the *STM (SHOOT MERISTEMLESS)* activity (Long & Barton 1998).

CLAVATA1 is a leucine-rich receptor-like kinase. The *Arabidopsis thaliana* genome project has so far identified over 50 LRR transmembrane receptor serine/threonine kinases. It can be estimated that the total *Arabidopsis thaliana* genome contains in the order of 100 family members. Other members of the *CLAVATA* family are *ERECTA* (*ER*) (Torii et al. 1996), *BRASSINOLIDE INSENSITVE* (*BRI*) (Li & Chory 1997) and *HAESA* (*HAE*) (Jinn, Stone, & Walker 2000), genes which have a function in plant development. Analogues of the *Arabidopsis thaliana CLAVATA* genes were also found in soybean (Glycine max). They were isolated from the wild type and the fasciating mutant of soybean and were designated GmCLVA1 and GmCLV1B (Yamamoto, Karakaya, & Knap 2000).

It is the aim of the present invention to isolate novel promoters and genes which are useful for altering features of transgenic plants and thus confer selective advantage to transgenic plants.

All the aims of the present invention have been met by the formulation of the following preferred embodiments of the current invention.

SUMMARY OF THE INVENTION

In the present invention a DNA operon was isolated from *Arabidopsis thaliana*, which comprises a novel root-specific promoter, driving a gene encoding a Leucine Rich Repeat (LRR) receptor-like kinase. Unexpectedly, the inventors discovered that the new operon was specifically expressed in roots and more surprisingly in the meristem of the main and lateral roots, in the main and lateral root vascular tissue and in the lateral root primordia.

The surprisingly specific expression pattern, as well as the very strong expression level, makes the newly found promoter a very attractive and selective tool to drive root-specific expression of any gene of interest. The ability to target meristem cells of the root makes this promoter clearly distinct from the root-specific promoters presently known in the art. Another surprising aspect of the invention is the identification of the new LRR receptor-

like kinase gene, which is the first gene identified in roots showing homology with a clavata-gene. This gene is specifically expressed in roots in a specific developmental stage of the root cells (namely in the meristematic phase) and can therefore contribute agronomic interesting features to a transgenic plant when transfected herein. Therefore the root-specific promoter is further denominated as ROOT *CLAVATA* HOMOLOG 1 promoter (RCH1prom) and the LRR receptor-like kinase as RCH1.

Accordingly, the invention embodies an isolated DNA sequence with nucleotide sequence as given in SEQ ID NO 1 (Figure 5) comprising a novel root specific regulatory promoter sequence with nucleotide sequence as given in SEQ ID NO 2 (Figure 6) or SEQ ID NO 18 (Figure 13), and an isolated DNA sequence with nucleotide sequence as given in SEQ ID NO 3 (Figure 8), encoding a novel ROOT *CLAVATA* HOMOLOG 1 gene, with amino acid sequence given in SEQ ID NO 4 or SEQ ID NO 19 (Figure 9).

Furthermore the invention embodies a method for modifying cell fate and/or plant development and/or plant morphology and/or biochemistry and/or physiology comprising the modification of expression in particular cells of the root of a particular transgene operably linked to the RCH1 promoter sequence of the present invention. Provided in the current invention are methods to effect expression of any gene of interest in particular cells of the plant root. The present invention also relates to a method to effect expression of a ROOT *CLAVATA* HOMOLOG 1 protein or a homologue or a derivative thereof in a plant cell, tissue or organ. The present invention clearly extends to any plant produced by the inventive methods described herein.

DETAILED DESCRIPTION OF THE INVENTION

One of the technical problems underlying the present invention is to provide regulatory elements that are active in the root only and that show no expression in other tissues. Additional to the highly selective and tissue-specific regulation of a promoter, other features such as very strong activity, continuous activity, and termination of activity at a certain developmental stage of the cells, would contribute an unique combination of regulatory elements to specifically regulate the expression of a gene of interest in the roots of a transgenic plant, thereby providing a means for modifying cell fate and/or plant development and/or plant morphology and/or biochemistry and/or physiology. Also the isolation of root-specific DNA sequences encoding proteins that could play a role in root

development is envisaged for use in modification of cell fate and/or plant development and/or plant morphology and/or plant biochemistry and/or lant physiology.

In a first embodiment, the present invention relates to an isolated nucleic acid comprising a transcriptional regulatory root promoter sequence and/or a nucleic acid sequence encoding a novel LRR receptor-like kinase (protein) or an immunologically active and/or functional fragment of such a regulary element or protein (kinase) selected from the group consisting of:

(a) nucleic acid comprising at least part of the DNA sequence as given in any of SEQ ID NOs 1, 2, 3 or 18, or the complement thereof,

(b) nucleic acid comprising the RNA sequences corresponding to at least part of any of SEQ ID NOs 1, 2, 3 or 18, or the complement thereof,

(c) nucleic acid specifically hybridizing, preferably under stringent conditions, with the nucleotide sequence as defined in (a) or (b),

(d) nucleic acid encoding a protein comprising the amino acid sequence as given in SEQ ID NO 4, SEQ ID NO 5 or SEQ ID NO 19,

(e) nucleic acid which is degenerated as a result of the genetic code to a nucleotide sequence encoding a protein as given in SEQ ID NO 4, SEQ ID NO 19 or to a nucleotide sequence as defined in (a) to (d),

(f) nucleic acid which is diverging due to the differences in codon usage between the organisms to a nucleotide sequence encoding a protein as given in SEQ ID NO 4, SEQ ID NO 19 or to a nucleic acid sequence as defined in (a) to (d),

(g) nucleic acid encoding a protein as given in SEQ ID NO 4, SEQ ID NO 19 or as defined in (a) to (d), which is diverging due to the differences between alleles encoding a protein

(h) nucleic acid encoding a fragment of a protein encoded by a DNA sequence as given in SEQ ID NO 1 or SEQ ID NO 3 or encoding a fragment of a protein encoded by a nucleic acid as defined in any one of (a) to (g), wherein said fragment comprises the sequence as represented in SEQ ID NO 5, and,

(k) nucleic acid encoding a protein as defined in SEQ ID NO 4, SEQ ID NO 19 or nucleic acid as defined in any one of (a) to (h), said nucleic acid interrupted by intervening DNA sequences.

The sequence as depicted in SEQ ID NO 1 is part of a sequence deposited under the Genbank Accession number AB017061, which represents the entire *Arabidopsis*

thaliana chromosome 5. The nucleic acid of the invention is not the nucleic acid with the sequence as deposited under the Genbank Accession number AB017061 (Figure 14).

Furthermore, in a preferred embodiment, the present invention relates to an isolated nucleic acid comprising a novel root-specific promoter sequence, termed RCH1prom, isolated from *Arabidopsis thaliana* and consisting of the nucleic acid as represented in SEQ ID NO 2 or SEQ ID NO 18, or a functional part of said sequence, which is able to regulate gene expression in a root specific manner as SEQ ID NO 2 or SEQ ID NO 18 itself. More specifically, said isolated DNA sequence provides a novel type of a root-specific promoter, which differs from other known root-specific promoters in the fact that it is unexpectedly active in only distinct cells of the root, namely the meristem of the main and the lateral root, as well as in the vascular tissue of the main and lateral root and in the lateral root primordia.

The isolation of the root specific promoter RCH1 of the present invention is described in Example 1.

Another aspect of the invention is the isolation and sequencing of a root-specific Leucin Rich Repeat (LRR) receptor-like kinase gene from root tissue of *Arabidopsis thaliana* and is also described in Example 1. The nucleotide sequence is listed in the current specification as SEQ ID NO 3 and illustrated in Figure 8. This sequence is present as part of the sequence deposited under the GenBank Accession number AB017061, which represents the entire *Arabidopsis thaliana* chromosome 5.

The novelty of the RCH1 promoter sequence relative to known root-specific promoters is substantiated by its unique expression pattern in plant roots. The analysis of the RCH1 expression patterns was done using a RCH1prom-GUS hybrid operon that was transformed to *Arabidopsis thaliana* as outlined in Example 2. Analogous experiments with the RCH1prom-GUS hybrid operon were performed to establish the expression pattern of RCH1 in embryo's (Example 8 and Figure 15).

The RCH1 promoter is active in the meristem of the main and the lateral roots, as well as in the vascular tissue of the main and the lateral root and in the lateral root primordia. The promoter of the present invention is the first known promoter that regulates a *CLAVATA* HOMOLOG gene in roots, and is clearly distinct from other described *CLAVATA*-operons which are all active in the shoot and floral meristem. The present data thus clearly distinguish RCH1prom from other plant root-specific promoters, which

are not expressed in the meristem of the main and lateral root, in the vascular tissue of the main and lateral root and in the lateral root primordia.

From the results obtained by the RCH1::GFP expression studies in root it was even more clear that the RCH1 promoter is strongly active in the endodermis, cortex, epidermis, lateral root cap, in short in the "division zone". The RCH1 promoter is also active but in a lower manner, in the quiescent center and in the vascular tissue.

A further characteristic of the RCH1 promoter is its very high expression level. Based on the incubation time of the RCH1prom-GUS hybrid operon transformed plant with the GUS substrate (1 hour) and the very high output signal in the root of that plant (Example 2, Figure 7D), it is clear for persons skilled in the art that the RCH1 promoter is very strong.

A further characteristic of the present invention is that the activity of the RCH1 promoter in the meristem cells stops when these cells start to differentiate and to elongate to become cells of specified root tissue other than root vascular tissue. Surprisingly the promoter is activated again in cells of the pericycle that form lateral root primordia. This expression is extended in the meristem of the lateral root, but again, the cells that differentiate into lateral root tissue other than vascular tissue do not retain RCH1 promoter activity.

Functional parts of the promoter of the invention will have at least one of these characteristics.

According to a further embodiment, the present invention relates to an isolated nucleic acid comprising a transcriptional regulatory root promoter comprising at least part of a sequence as given in SEQ ID NO 1, 2 or 18.

According to another embodiment, the present invention relates to an isolated nucleic acid as represented in SEQ ID NO 2 or 18 wherein said transcriptional regulatory root promoter is operably linked to said nucleic acid sequence encoding a novel LRR receptor-like kinase or to an immunologically active and/or functional fragment thereof.

According to the invention, distinctive fragments of the newly identified promoter, which contribute specific features to said promoter, can be fused to another nucleic acid sequence (e.g. a promoter) in order to alter the sequential and/or regulatory features of the latter.

The invention thus relates to an isolated nucleic acid having transcriptional regulatory root specific, root-meristem specific or root-vascular-tissue specific promoter activity

comprising at least part of the DNA sequence as given in any of SEQ ID 1, 2 or 18 and a second transcriptional regulatory sequence.

In one embodiment, said second trancriptional regulatory sequence is for instance an inducible box or sequence, which is placed within the RCH1 promoter, wherein the function of the RCH1 promoter is not hampered but instead can be regulated, for instance by induction. In the latter embodiment, the RCH1 promoter is not used to confer root-specificity to other promoters, but a second sequence is used to confer an additional regulatory feature to the RCH1 promoter.

In another embodiment said second transcriptional regulatory sequence is a promoter sequence not normally exhibiting root-specificity, thereby resulting in a hybrid promoter with root-specificity.

Accordingly, another embodiment of the present invention is to confer root-specificity, root-meristem-specificity, and root-vascular-tissue-specificity and/or root abundant gene expression to other promoter sequences. This can be achieved by fusing elements of the here-disclosed RCH1 promoter to the promoter sequence of interest preferably the widely used ubiquitin promoter. Such modifications can be achieved by routine experimentation by those skilled in the art.

It should be further understood that the invention preferably relates to any of the isolated nucleic acids as described above which is DNA, cDNA, genomic DNA, synthetic DNA or RNA wherein T is replaced by U.

The invention also relates to a nucleic acid molecule of at least 15 nucleotides in length hybridizing specifically with any of the nucleic acids of the invention. The invention also relates to a nucleic acid molecule of at least 15 nucleotides in length specifically amplifying a nucleic acid of the invention.

Further, the invention relates to a vector comprising a nucleic acid of the invention. Said vector may be an expression vector wherein the (said) nucleic acid sequence encoding a novel LRR receptor-like kinase or an immunologically active and/or functional fragment thereof, is operably linked to one or more control sequences allowing the expression in prokaryotic and/or eukaryotic host cells.

The invention also relates to a vector comprising at least part of a nucleic acid according to the invention and wherein a transcriptional regulatory root promoter of the invention or functional part(s) thereof, is operably linked to one or more genes of interest.

The invention also relates to a host cell containing a nucleic acid molecule of the invention or a vector of the invention wherein said nucleic acid or vector has been introduced by transfomation, transfection or infection. Preferred host cells that are part of the invention are bacterial, insect, fungal, plant or animal cells.

Therefore, in a preferred embodiment, the invention relates to an isolated nucleic acid as represented in SEQ ID NO 3 encoding a LRR receptor-like kinase characterized by an amino acid sequence as given in SEQ ID NO 4 or 19. More specifically, said isolated nucleic acid encodes a plant ROOT CLAVATA HOMOLOG 1 (RCH1) of a novel type which is unexpectedly expressed in roots. Said novel plant RCH1 shows significant homology with CLAVATA 1, a protein involved in the maintenance of floral and shoot meristem cells.

An initial comparative amino acid sequence homology search using the BLASTP 2.0.8 software (Altschul et al. 1997) revealed that most significant alignments of RCH1 protein are produced with receptor-like protein kinases, more specifically with leucine rich receptor-like kinases such as CLAVATA1 (CLV1). Furthermore, the most homologous protein with a known function of this list, is the CLAVATA 1 receptor-like kinase from *Arabidopsis thaliana.* Therefore the novel gene was denominated a ROOT *CLAVATA* HOMOLOG 1 gene (RCH1).

An amino acid sequence alignment of the RCH1 protein and *CLAVATA* 1 from *A. thaliana* is shown in Figure 10. Calculation of the percentage of identical amino acids residues between RCH1 and CLV1 teaches that RCH1 is 39% identical to CLV1. Additional to 39% of identical amino acids, there is 49% of the amino acids that belong to the same class of amino acids and therefore have the same physical and chemical properties.

The classification of RCH1 as a novel type of *CLAVATA* is documented further in figures 11 and 12.

The conclusions we can draw from figure 11 is that based on the sequence, RCH1 is closer related to CLV1 compared to other known CLV1 homologues, such as ERECTA, BRASSINOLIDE INSENSITIVE and HAESA.

The fact that the novel LRR receptor like protein described in the present invention is rather a root clavata 1 homologue then another kind of LRR receptor-like kinase (e.g. resistance genes) is illustrated in Figure 12 in which the evolutionary relationship between clavata homologues and resistance genes is shown. The resistance genes

(RPS5, RPS2, RPP8, RPM1, RPS4, N, RPP5) are described in "Meyers BC, Dickerman AW, Michelmore RW, Sivaramakrishnan S, Sobral BW, Young ND. Plant disease resistance genes encode members of an ancient and diverse protein family within the nucleotide-binding super family. Plant J. 1999 Nov;20(3):317-32".

The present invention relates to a ROOT *CLAVATA* HOMOLOG 1 gene that differs from other known *CLAVATA*-like genes, in the fact that it is active in the root meristem, whereas the other known *CLAVATA* genes and CLAVATA homologs (like ERECTA, BRASSINOLIDE INSENSITIVE and HAESA) are only active in the shoot and floral meristem.

According to another embodiment, the invention also relates to nucleic acids encoding a functional plant LRR receptor-like kinase comprising one or more protein regions, distinguishing said plant LRR receptor-like kinase from those LRR receptor-like kinases known in the art.

The inventors found that the predicted amino acid sequence (BAB10317.1, receptor protein kinase-like protein) corresponding to the database entry with accession number AB017061 (A. thalinana chromosome 5), was missing a part of the amino acid sequence (namely, 25 amino acids as represented in SEQ ID NO. 5 of the present patent application). This extra stretch of 25 amino acids was identified by the cloning and the sequencing of the corresponding cDNA of the present invention. Therefore, the predicted prior art sequence is believed not to represent a functional enzyme.

The invention thus relates to an isolated nucleic acid selected from the group consisting of:

(a) a nucleic acid consisting of at least part, preferably at least a functional part, of the DNA sequence as given in SEQ ID NO 1 or 3, or the complement thereof,

(b) a nucleic acid encoding a protein as given in SEQ ID NO 4 or 19 or encoding a fragment of said protein, wherein said fragment comprises the sequence as represented in SEQ ID NO 5, and

(c) a nucleic acid encoding a protein with an amino acid sequence which is at least 60%, preferably at least 65 %, 70 %, 75%, 80%, 85%, more preferably at least 90% or 95% identical to the protein as given in SEQ ID NO 4 or 9, wherein said amino acid sequence comprises the sequence as represented in SEQ ID NO 5,

characterised in that said nucleic acid encodes a novel LRR receptor-like kinase protein or an immunologically active and/or functional fragment of such a protein, and further provided that said nucleic acid is not one of the nucleic acids as deposited under the GenBank Accession numbers AB017061 or AQ966419.

The present invention also relates to an isolated LRR receptor-like kinase (protein) comprising one of the polypeptides selected from the group consisting of:

(a) a polypeptide as given in SEQ ID NO 4 or SEQ ID NO 19,

(b) a polypeptide comprising the amino acid sequence as given in SEQ ID NO 5, and,

(c) a polypeptide comprising the sequence represented in SEQ ID NO 5 and encoded by a nucleic acid as given in SEQ ID NO 1 or 3,

or a homologue or a derivative of said protein, or a fragment or an immunologically active and/or functional fragment thereof, provided that said homologue is not the amino acid sequence as described under the GenBank Accession number BAB10317.1.

With the expression "homologue" when referring to homologues of the LRR receptor-like kinase (protein), a polypeptide is meant which is at least 65% identical to the amino acid sequence of RCH1 as represented in SEQ ID NO 4 or 19.

More preferred, the invention relates to a protein consisting of an amino acid sequence as given in SEQ ID NO 4 or SEQ ID NO 19. More specifically, the present invention relates to a root CLAVATA homologue or a functional homologue thereof. A preferred fragment of said protein is represented in SEQ ID NO 5.

A further embodiment of the invention comprises homologues, derivatives, immunologically active and/or functional fragments of the LRR receptor-like kinase (protein) according to the invention, and proteins comprising said homologues, derivatives, immunologically active and/or functional fragments of said LRR receptor like kinase (protein).

As such, the present invention also relates to an isolated polypeptide comprising the sequence represented in SEQ ID NO 5 encodable by a nucleic acid molecule of the invention as defined above, or a homologue or a derivative thereof, or a fragment or an immunologically active and/or functional fragment thereof.

In a more preferred embodiment, the invention relates to a polypeptide, encodable by a nucleic acid molecule of the invention and which has an amino acid sequence as given in SEQ ID NO 4 or SEQ ID NO 19 or comprising an amino acid sequence as given in

SEQ ID NO 5, or a homologue or a derivative thereof, or an immunologically active and/or functional fragment thereof.

Any of said proteins can be produced in a biological system, e.g. a cell culture. Alternatively any of said proteins is chemically manufactured e.g. by solid phase peptide synthesis. Said proteins or fragments thereof can be part of a fusion protein as is the case in e.g. a two-hybrid assay which enables e.g. the identification of proteins interacting with the LRR receptor-like kinase according to the present invention.

Therefore, according to another embodiment, the invention also relates to a method for producing a polypeptide or a protein of the invention comprising culturing a host cell further specified above under conditions allowing the expression of the polypeptide and recovering the produced polypeptide from the culture.

The proteins or fragments thereof obtained by a method of the invention are furthermore useful e.g. to modulate the interaction between a LRR receptor-like kinase according to the invention and a ligand of the receptor and/or other identified interacting protein partners. Chemically synthesized peptides are particularly useful e.g. as a source of antigens for the production of antisera and/or antibodies.

The current invention thus furthermore encompasses antisera and/or antibodies specifically recognizing the LRR receptor-like kinase according to the invention or immunologically active parts or epitopes thereof. Said antisera and/or antibodies are useful in many areas related to the invention including: (i) identification in any organism, preferably plants, of other LRR receptor-like kinase and their genes according to the invention; (ii) quantification of synthesis in organisms and/or recombinant organisms of the LRR receptor-like kinase according to the invention; (iii) purification of the LRR receptor-like kinase according to the invention; (iv) immunoprecipitation of the LRR receptor-like kinase according to the invention e.g. as a way to identify other protein partners complexing with said LRR receptor-like kinase; (v) immunolocalization of the LRR receptor-like kinase according to the invention which is expressed in an organism or a recombinant organism.

The invention also relates to a method for the production of transgenic plants, plant cells or plant tissues comprising the introduction of a nucleic acid molecule of the invention in an expressible format or a vector as described earlier in said plant, plant cell or plant tissue

The invention also relates to a method as described above further comprising regenerating a plant from said plant cell.

In a further embodiment, the invention also relates to a transgenic plant cell comprising a nucleic acid sequence of the invention which is operably linked to regulatory elements allowing transcription and/or expression of said nucleic acid in plant cells or obtainable by one of the methods described above. Preferably, in said transgenic plant cell one of the nucleic acids of the invention is stably integrated into the genome of said plant cell.

The invention further relates to a transgenic plant or plant tissue comprising said plant cells and to a harvestable part of said plant or tissue. Said harvestable part of said plant or tissue is preferably selected from the group consisting of seeds, leaves, fruits, stem cultures, rhizomes, tubers and bulbs.

The invention further extends to the progeny derived from any of the plants or plant parts described above.

The present invention is applicable to any plant, in particular a monocotyledonous plants and dicotyledonous plants including a fodder or forage legume, ornamental plant, food crop, tree, or shrub selected from the list comprising *Acacia spp., Acer spp., Actinidia spp.,Aesculus spp., Agathis australis, Albizia amara, Alsophila tricolor, Andropogon spp., Arachis spp, Areca catechu, Astelia fragrans, Astragalus cicer, Baikiaea plurijuga, Betula spp., Brassica spp., Bruguiera gymnorrhiza, Burkea africana, Butea frondosa, Cadaba farinosa, Calliandra spp, Camellia sinensis, Canna indica, Capsicum spp., Cassia spp., Centroema pubescens, Chaenomeles spp.,Cinnamomum cassia, Coffea arabica, Colophospermum mopane, Coronillia varia, Cotoneaster serotina, Crataegus spp., Cucumis spp., Cupressus spp., Cyathea dealbata, Cydonia oblonga, Cryptomeria japonica, Cymbopogon spp., Cynthea dealbata, Cydonia oblonga, Dalbergia monetaria, Davallia divaricata, Desmodium spp., Dicksonia squarosa, Diheteropogon amplectens, Dioclea spp, Dolichos spp., Dorycnium rectum, Echinochloa pyramidalis, Ehrartia spp., Eleusine coracana, Eragrestis spp., Erythrina spp., Eucalyptus spp., Euclea schimperi, Eulalia villosa, Fagopyrum spp., Feijoa sellowiana, Fragaria spp., Flemingia spp, Freycinetia banksii, Geranium thunbergii, Ginkgo biloba, Glycine javanica, Gliricidia spp, Gossypium hirsutum, Grevillea spp., Guibourtia coleosperma, Hedysarum spp., Hemarthia altissima, Heteropogon contortus, Hordeum vulgare, Hyparrhenia rufa, Hypericum erectum, Hyperthelia dissoluta, Indigo incarnata, Iris spp., Leptarrhena pyrolifolia, Lespediza spp., Lettuca spp., Leucaena leucocephala, Loudetia simplex,*

Lotonus bainesii, Lotus spp., Macrotyloma axillare, Malus spp., Manihot esculenta, Medicago sativa, Metasequoia glyptostroboides, Musa sapientum, Nicotianum spp., Onobrychis spp., Ornithopus spp., Oryza spp., Peltophorum africanum, Pennisetum spp., Persea gratissima, Petunia spp., Phaseolus spp., Phoenix canariensis, Phormium cookianum, Photinia spp., Picea glauca, Pinus spp., Pisum sativum, Podocarpus totara, Pogonarthria fleckii, Pogonarthria squarrosa, Populus spp., Prosopis cineraria, Pseudotsuga menziesii, Pterolobium stellatum, Pyrus communis, Quercus spp., Rhaphiolepsis umbellata, Rhopalostylis sapida, Rhus natalensis, Ribes grossularia, Ribes spp., Robinia pseudoacacia, Rosa spp., Rubus spp., Salix spp., Schyzachyrium sanguineum, Sciadopitys verticillata, Sequoia sempervirens, Sequoiadendron giganteum, Sorghum bicolor, Spinacia spp., Sporobolus fimbriatus, Stiburus alopecuroides, Stylosanthos humilis, Tadehagi spp, Taxodium distichum, Themeda triandra, Trifolium spp., Triticum spp., Tsuga heterophylla, Vaccinium spp., Vicia spp.Vitis vinifera, Watsonia pyramidata, Zantedeschia aethiopica, Zea mays, amaranth, artichoke, asparagus, broccoli, brussel sprout, cabbage, canola, carrot, cauliflower, celery, collard greens, flax, kale, lentil, oilseed rape, okra, onion, potato, rice, soybean, straw, sugarbeet, sugar cane, sunflower, tomato, squash, and tea, amongst others, or the seeds of any plant specifically named above or a tissue, cell or organ culture of any of the above species.

Yet according to another embodiment the present invention relates to a method for conferring root-specificity and/or root-meristem-specificity and/or root-vascular-tissue-specificity and/or root-endodermis-specificity and/or root-cortex-specificity and/or root-epidermis-specificity and/or lateral-root cap-specificity and/or abundant expression in roots to other promoter sequences comprising the fusion of at least part of the DNA sequence as given in SEQ ID NO 1, 2 or 18 to a (second) transcriptional regulatory promoter sequence normally not exhibiting root-specificity.

The invention further relates to method for root-specific expression of a gene(s) of interest comprising operably linking of said gene(s) of interest to a transcriptional regulatory root-specific promoter comprising at least part of the nucleic acids as given in SEQ ID NO 1, 2 or 18 as defined above, and possibly in combination/or a (second) transcriptional regulatory promoter sequence not normally exhibiting root-specificity, as defined earlier.

The invention further relates to a method for modifying cell fate and/or plant development and/or plant morphology and/or biochemistry and/or physiology comprising the modification of expression in the meristem of the main and the lateral roots, or in the vascular tissue of the root, or in the lateral root primordia of a gene(s) of interest operably linked to a transcriptional regulatory root promoter as defined earlier.

The ability to drive tissue-specific expression of any gene of interest in plants is considered to be of great agronomic importance. This art can be used to exclude expression in eatable parts of the plant, which averts the uptake of ectopic components. Alternatively the expression of the transgene can be restricted to the relevant tissues of the plant so the total organism does not spend too much energy in superfluous production of the foreign component. This technique can also reduce potential yield loss by limiting the expression of some pernicious, yet useful agronomic genes, to specific tissues only. Said genes of interest, which can be expressed in a root-specific manner under the control of the present invention can also be involved in phytoremediation, enhanced root-growth (like cell cycle genes or cytokinin oxidase) or disease tolerance. These methods are more elaborated in detail below.

The invention described here provides a new transcriptional regulatory element which can be operably linked to a gene effecting and/or modifying root metabolites which can be involved in phytoremediation. Examples are genes involved in the production of citric and maleic acids that can be secreted by the plant and subsequently form complexes with the aluminium in the soil. Root specific expression of the aluminium-resistance gene that give rise to said metabolites can therefore allow plants to grow on acid soils, like those of south-eastern United States, Central and south America, North Africa and parts of India and China, where the aluminium is set free as an ion that poisons plant roots.

Alternatively, the promoter of the present invention also provides a tool to be used in a strategy for environmental remediation. Accordingly, a metal-resistance gene such as the TaPCS1 or the CAD1 gene, contributing cadmium tolerance, can be expressed specifically in roots to enable enhanced growth of transgenic plants on soils contaminated with cadmium. The metals can consequently be extracted from the soil and stored in the plant material, which could then be harvested and incinerated.

Therefore according to preferred embodiments, the present invention relates to a method for phytoremidiation or environmental remediation comprising the expression of

a gene(s) of interest under the control of a transcriptional regulatory root promoter as defined earlier.

In conclusion the present invention provides a method for broadening the range of soils or environments in which a plant may thrive or that can be used for agriculture.

Plant root development is an essential determinant of plant growth and crop yield, so that yield could be enhanced by induced changes in expression of root-specific regulatory factors. Cyclin D for example, which promotes the cell division, can be expressed under the root-specific control of the current invention and hence contribute to increased growth rate and crop yield. Also genes that are involved in biosynthesis and perception of phytohormones like auxins, cytokinines, ethylene, abscisic acid and gibeberellins can be used to enhance root growth and development. However, the root-specific expression of these genes might be pivotal for the overall welfare of the transgenic plant.

It is known that the constitutive and ubiquitous expression of the cytokinine oxidase in *Arabidopsis thaliana* leads to the promotion of an enlarged root system and enhanced root biomass, but also to diminished shoot growth. To overcome this problem the cytokinin oxidase can be operably linked to the promoter of the present invention as described in example 6, which will enhance root growth without negatively affecting shoot growth. This enhanced root growth in the plant can be used for bioremidiation or phytoremediation.

Also the said gene of interest, that can be operably linked to the promoter of the present invention, can be a root transcription factor, such as Alfin. Alfin binds to promoter fragments of predominantly root-specific and salt inducible genes and its activity has major implications for crop plant yield (Winicov 2000) (Winicov & Bastola 1999)). In conclusion the present invention provides a method for enhancing plant growth and crop yield.

It is also clear for the skilled artisan that the promoter of the present invention provides a useful tool to be used for the introduction of resistance to soil-borne pathogen attack to plants, a widespread target for genetic improvement of crop plants. Certain disease-inducing microorganisms attack the belowground plant tissue and any genetic modification to contribute resistance to such organisms will require expression of the resistance-gene in the roots. To overcome this problem, the coding region that is operably linked to the promoter of the present invention preferably encodes a protein

which is toxic to root-attacking organisms and more preferably the protein is an insecticidal endotoxin of *Bacillus thuringiensis*. Similarly to the above-mentioned pathogens, there are pests attacking plants. Certain insect species attack green leaf tissue, whereas other, for example coleoptera, attack the roots. Corn root worm (*Diabrotica undecimpunctata howardi Barber*) for example is a particularly difficult pest to control or to eradicate. It attacks the plant below the soil line, where insecticides are difficult or impossible to apply effectively. Corn root worm can be eradicated by root-specific expression of the Cry III gene, which produces a component, specifically toxic for coleoptera. Also root-specific expression of the limonene cyclase gene in combination with GPPP synthetase would be larvicidal. Another pest, nematodes, that penetrate the root and form root-knots or cysts, which cause severe damage to crops throughout the world, could be treated with the root-specific expression of nematode resistance genes such as $Hs1^{pro-1}$, Ni-1 and Gpa2.

In conclusion the invention disclosed here provides a method for conferring enhanced resistance to pathogens to a transgenic plant which pathogens attack the belowground plant tissue comprising the expression of a gene(s) of interest under the control of a transcriptional regulatory root-specific promoter as defined earlier.

In yet another embodiment of the invention, the invention provides a method to enhanced freezing tolerance comprising the expression of a gene(s) of interest under the control of a transcriptional regulatory root-specific promoter as defined earlier. Preferred is the overexpressing of the *Arabidopsis thaliana* DNA binding factors CBF1 and DREB1A that induce cold-regulated (COR) genes (Jaglo-Ottosen et al. 1998). However, constitutive expression of these genes in all organs of the plant led to growth retardation under normal growth conditions. Root-specific expression of these freezing-tolerance genes under control of the promoter sequence of the present invention can make the plant tolerant to cold temperatures of the soil without affecting the growth of the plant. This technique can have important agronomic potential since lots of soils are inaccessible for crop production because of presence of permanently frozen layers in the soil.

Alternatively, other genes of interest can tolerize the transgenic plant for other weather conditions like drought (e.g. oat arginin decarboxylase gene) and high salt conditions (e.g. MtID gene). In conclusion, the present invention provides a method to contribute to a transgenic plant stress-tolerance.

Further advantages to root-preferred gene expression include the production of useful proteins in an industrial setting. Light-sensitive proteins may be synthesized in root tissue such that said proteins are not exposed to light. Therefore the present invention also relates to a method for the production of light-sensitive proteins comprising the expression of a gene encoding said light-sensitive protein under the control of a transcriptional regulatory root-specific promoter of the invention.

Also considered as an embodiment of the invention is the use of the RCH1 promoter for applications based on the inhibition of expression of native DNA sequences within the plant's root to achieve a desired phenotype resulting from the silencing of the native gene. In this case, such inhibition might be accomplished with transformation of the plant to comprise a transcriptional regulatory root-specific promoter of the present invention operably linked to e.g. an antisense nucleotide sequence to the gene of interest, a gene silencing construct or a ribozyme.

Those skilled in the art will be aware that plants secure the formation of organs throughout their life span by developing and maintaining a collection of stem cells termed the meristem. The promoter of the present invention offers the opportunity to deliver a certain protein specifically to these cells. For example overexpression of genes that influence the proliferation and/or the differentiation of the root meristem could contribute to a higher turnover of the meristem cells and therefor enhance the root architecture.

The promoter of the present invention drives the *Arabidopsis thaliana* gene ROOT *CLAVATA* HOMOLOG 1 (RCH1). Since the *clavata*-genes are key role players in shoot-meristem maintenance, persons skilled in the art may recognize the possible involvement of the RCH1 operon in root meristem maintenance. Therefor one could assume that the control of this promoter is influenced by signals that promote meristem maintenance. This might be an advantage when aiming to stimulate the root meristem-formation, which eventually could result in more roots, higher uptake of resources and in higher crop yield.

Therefore, according to preferred embodiments, the invention relates to methods for stimulating meristem formation and/or maintaining root meristem comprising the expression of a gene that influences the proliferation and/or differentiation of the root mersiteme under the control of a transcriptional regulatory root (-specific) promoter of the invention.

The present invention therefore relates to a method for stimulating root meristem formation or for root-meristeme maintenance comprising the expression in particular cells, tissues or organs of a plant, of a nucleic acid encoding an LRR receptor-like kinase, for instance the LRR receptor-like kinase of the invention, wherein optionally said nucleic acid is operably linked to a plant-operable promoter sequence, for instance said plant-operable promoter being the transcriptional regulatory root promoter as defined earlier.

As mentioned supra, the combined gene expression data and homology to CLAVATA 1 indicate that the novel plant LRR receptor-like kinase of the invention, RCH1 (SEQ ID No. 3, 4, 19), is very likely to exert a function that is involved in the development and/or maintenance of the root meristem. As the root is the main plant organ for nutrient uptake from the environment, enhanced root formation could eventually lead to overall growth enhancement and better yield. Therefore preferred genes to be expressed in the methods for stimulating meristem formation and/or for root meristeme maintenance are genes encoding the polypeptide(s) of the invention.

The present invention also relates to a method for enhancing root formation and/or root growth comprising the expression in particular cells, tissues or organs of a plant, of a nucleic acid encoding an LRR receptor-like kinase, for instance the LRR receptor-like kinase of the invention, wherein optionally said nucleic acid is operably linked to a plant-operable promoter sequence, for instance said plant-operable promoter being the transcriptional regulatory root promoter as defined earlier.

The present invention therefore relates to a method for enhancing overall growth and yield comprising the expression in particular cells, tissues or organs of a plant, of a nucleic acid encoding an LRR receptor-like kinase, for instance the LRR receptor-like kinase of the invention, wherein optionally said nucleic acid is operably linked to a plant-operable promoter sequence, for instance said plant-operable promoter being the transcriptional regulatory root promoter as defined earlier.

The present invention also relates to a method for enhancing overall growth and yield comprising the expression of a protein of the invention under the control of a transcriptional regulatory root (or root-specific) promoter of the invention.

Also in yet other embodiments of the invention, methods are provided for modifying cell fate and/or plant development and/or plant morphology and/or biochemistry and/or

physiology comprising the modification of expression in particular cells, tissues or organs of a plant, of a nucleic acid encoding an LRR receptor-like kinase, for instance the LRR receptor-like kinase of the invention and defined above, wherein optionally said nucleic acid is operably linked to a plant-operable promoter sequence, for instance said plant-operable promoter being the transcriptional regulatory root promoter as defined earlier.

The present invention further relates to a method to confer pathogen resistance to a transgenic plant comprising the expression in particular cells, tissues or organs of a plant, of a nucleic acid encoding an LRR receptor-like kinase, for instance the LRR receptor-like kinase of the invention and defined above, wherein optionally said nucleic acid is operably linked to a plant-operable promoter sequence, for instance said plant-operable promoter being the transcriptional regulatory root promoter as defined earlier.
Many pathogen resistance genes are LRR receptor-like proteins, which initiate a signal transduction pathway in the root cells, leading to the activation of defense mechanisms against the pathogen. The novel LRR receptor-like kinase might contribute to pathogen resistance to the transgenic plant when transformed herein.
Yet another embodiment of the invention relates to methods to perform a functional analysis of the DNA sequence according to the present invention, encoding a novel LRR receptor-like protein. For persons skilled in the art it is possible to make single or double knock-out mutants of *Arabidopsis thaliana*, wherein the DNA of the present invention is inactivated. Also introduction of the antisense sequence of the DNA of the invention, as well as the overexpression of its sense sequence in a transgenic plant can elucidate a specific feature in this transgenic plant. Furthermore, the loss of function of this gene of the invention by introduction of dominant negative mutations can result in interesting characteristics of the mutant plant.
Another aspect of the present invention is the revelation of a new target for herbicides or growth regulators. When said gene of the invention is an essential gene for plant cell viability, it can be a target for a herbicide or growth regulator.
Therefore the invention relates to a method for identifying and obtaining proteins interacting with a (LRR receptor-like kinase) polypeptide of the invention comprising a screening assay wherein said polypeptide is used or expressed.
In a preferred embodiment the invention relates to a method for identifying and obtaining proteins interacting with a an LRR receptor-like kinase (protein) comprising a two-hybrid

screening system wherein a nucleic acid encoding a polypeptide of the invention as a bait and a cDNA library as prey are expressed.

The invention also relates to a method for modulating the interaction between an LRR receptor-like kinase (protein) of the invention and interacting proteins obtainable by a method as described above.

The invention further relates to a method for identifying and obtaining compounds interacting with an LRR receptor-like kinase (protein) comprising the steps of:

a) providing a two-hybrid screening system wherein a polypeptide or protein of the invention and a protein interacting with said (LRR receptor-like kinase) polypeptide or protein or an interacting protein obtainable by a method of claim as claimed above are expressed,

b) interacting said compound with the complex formed by the expressed proteins as defined in a),

c) detecting a second complex, wherein the presence of said second complex identifies a compound which specifically binds to one of said polypeptides or said second complex, and

d) identifying the compound.

The invention also relates to a method for identifying compounds or mixtures of compounds which specifically bind to a polypeptide of the invention, comprising:

a) combining a polypeptide of the invention with said compound or mixtures of compounds under conditions suitable to allow complex formation, and,

b) detecting complex formation, wherein the presence of a complex identifies a molecule which specifically binds said polypeptide.

The invention further relates to the use of a molecule identified by means of one of the methods described above as a plant growth regulator or herbicide.

The invention also relates to a method for the production of a plant growth regulator or herbicide composition comprising the steps of one of the methods described above and formulating the compounds obtained from said steps in a suitable form for the application in agriculture or plant cell or tissue culture.

The invention also relates to the use of any of the nucleic acid molecules, vectors, polypeptides or antibodies of described herein for modifying cell fate and/or plant development and/or plant morphology and/or plant biochemistry and/or plant physiology.

The invention also relates to a diagnostic composition comprising at least one of the nucleic acid molecules, vectors, polypeptides or antibodies of the invention.

It is further important to notice that CLAVATA 1 has many analogues in other plant species, so it is likely that also the promoter of RCH1, the RCH1 ORF or the complete RCH1 operon can be functional in those other plant species, and possibly exert the same specific activation pattern. This can be investigated by root-specific reporter gene expression in rice mediated by the *Arabidopsis thaliana* RCH1 promoter as described in example 5.

Definitions and elaborations to the embodiments

The terms "protein(s)", "peptide(s)" or "oligopeptide(s)", when used herein refer to amino acids in a polymeric form of any length. Said terms also include known amino acid modifications such as disulphide bond formation, cysteinylation, oxidation, glutathionylation, methylation, acetylation, farnesylation, biotinylation, stearoylation, formylation, lipoic acid addition, phosphorylation, sulphation, ubiquitination, myristoylation, palmitoylation, geranylgeranylation, cyclization (e.g. pyroglutamic acid formation), oxidation, deamidation, dehydration, glycosylation (e.g. pentoses, hexosamines, N-acetylhexosamines, deoxyhexoses, hexoses, sialic acid etc.), acylation and radiolabels (e.g. ^{125}I, ^{131}I, ^{35}S, ^{14}C, ^{32}P, ^{33}P, ^{3}H) as well as non-naturally occurring amino acid residues, L-amino acid residues and D-amino acid residues.

"Homologues" or "Homologs" of a protein of the invention are those peptides, oligopeptides, polypeptides, proteins and enzymes which contain amino acid substitutions, deletions and/or additions relative to the said protein with respect to which they are a homolog, without altering one or more of its functional properties, in particular without reducing the activity of the resulting. For example, a homolog of said protein will consist of a bioactive amino acid sequence variant of said protein. To produce such homologs, amino acids present in the said protein can be replaced by other amino acids having similar properties, for example hydrophobicity, hydrophilicity, hydrophobic moment, antigenicity, propensity to form or break α-helical structures or β-sheet structures, and so on. An overview of physical and chemical properties of amino acids is given in Table 1.

Substitutional variants of a protein of the invention are those in which at least one residue in said protein amino acid sequence has been removed and a different residue

inserted in its place. Amino acid substitutions are typically of single residues, but may be clustered depending upon functional constraints placed upon the polypeptide; insertions will usually be of the order of about 1-10 amino acid residues. and deletions will range from about 1-20 residues. Preferably, amino acid substitutions will comprise conservative amino acid substitutions, such as those described *supra*.

Table 1: Properties of naturally occurring amino acids.

Charge properties / hydrophobicity	Side group	Amino Acid
nonpolar	Aliphatic	Ala, Ile, Leu,
hydrophobic	Aliphatic, S-containing	Val
	Aromatic	Met
	Imino	Phe, Trp
		Pro
polar uncharged	Aliphatic	Gly
	Amide	Asn, Gln
	Aromatic	Tyr
	Hydroxyl	Ser, Thr
	Sulfhydryl	Cys
positively charged	Basic	Arg, His, Lys
negatively charged	Acidic	Asp, Gly

Insertional amino acid sequence variants of a protein of the invention are those in which one or more amino acid residues are introduced into a predetermined site in said protein. Insertions can comprise amino-terminal and/or carboxy-terminal fusions as well as intra-sequence insertions of single or multiple amino acids. Generally, insertions within the amino acid sequence will be smaller than amino or carboxyl terminal fusions, of the order of about 1 to 10 residues. Examples of amino- or carboxy-terminal fusion proteins or peptides include the binding domain or activation domain of a transcriptional activator as used in the yeast two-hybrid system, phage coat proteins, (histidine)$_6$-tag, glutathione S-transferase, protein A, maltose-binding protein, dihydrofolate reductase, Tag•100

epitope (EETARFQPGYRS), c-myc epitope (EQKLISEEDL), FLAG®-epitope (DYKDDDK), lacZ, CMP (calmodulin-binding peptide), HA epitope (YPYDVPDYA), protein C epitope (EDQVDPRLIDGK) and VSV epitope (YTDIEMNRLGK).

Deletional variants of a protein of the invention are characterised by the removal of one or more amino acids from the amino acid sequence of said protein.

Homologues of the peptides or polypeptides of the invention contain a number of amino acid substitutions, deletions and/or additions relative to said peptide or polypepthide such that said homologues are still substantially different from the peptides or polypeptides known in the art. The difference between proteins can be calculated in terms of % identidty according to methods well known in the art. Preferably, the homologues of the peptides or polypeptides of the present invention comprise the amino acid sequence represented in SEQ ID NO 5, optionally comprising minor amino acid substitutions such as those described in table 1. Other homologues of the polypeptides of the invention have an amino acid sequence which is at least 65 % identical to the sequence represented in SEQ ID NO 4 or 19.

Amino acid variants of a protein of the invention may readily be made using peptide synthetic techniques well known in the art, such as solid phase peptide synthesis and the like, or by recombinant DNA manipulations. The manipulation of DNA sequences to produce variant proteins which manifest as substitutional, insertional or deletional variants are well known in the art. For example, techniques for making substitution mutations at predetermined sites in DNA having known sequence are well known to those skilled in the art, such as by M13 mutagenesis, T7-Gen in vitro mutagenesis kit (USB, Cleveland, OH), QuickChange Site Directed mutagenesis kit (Stratagene, San Diego, CA), PCR-mediated site-directed mutagenesis or other site-directed mutagenesis protocols.

"Derivatives" of a protein of the invention are those peptides, oligopeptides, polypeptides, proteins and enzymes which comprise at least about five contiguous amino acid residues of said polypeptide but which retain the biological activity of said protein and which may further comprise additional naturally-occurring, altered glycosylated, acylated or non-naturally occurring amino acid residues compared to the amino acid sequence of a naturally-occurring form of said polypeptide. Alternatively or in addition, a derivative may comprise one or more non-amino acid substituents compared to the amino acid sequence of a naturally-occurring form of said polypeptide, for

example a reporter molecule or other ligand, covalently or non-covalently bound to the amino acid sequence such as, for example, a reporter molecule which is bound thereto to facilitate its detection. Preferably, the derivates of the proteins of the present invention comprise the amino acid acid sequence as represented in SEQ ID NO 5.

With "immunologically active" is meant that a molecule or specific fragments thereof such as epitopes or haptens are recognized by, i.e. bind to antibodies. Preferably, the immunologically active or fragments of the proteins of the present invention comprise the amino acid acid sequence as represented in SEQ ID NO 5.

With "functional fragment" or "functional homologue" is meant a protein which comprises at least about 5,10, 20, 50, 75, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1050, 1100, 1110, 1120, 1130, 1131, 1132, 1133 or 1134 contiguous amino acid residues of said polypeptide but which retains the biological activity of said protein. Preferably, the functional fragments or functional homologues of the polypeptide or proteins of the invention comprise the amino acid sequence as represented in SEQ ID NO 5 but wherein minor amino acid substitutions are allowed (see table 1).

In the context of the current invention are embodied homologs, derivatives and/or immunologically active fragments of any of the inventive LRR receptor-like kinase protein or polypeptide or homolog, derivative or fragment thereof as defined supra.

"Antibodies" include monoclonal, polyclonal, synthetic or heavy chain camel antibodies as well as fragments of antibodies such as Fab, Fv or scFv fragments. Monoclonal antibodies can be prepared by the techniques as described in e.g. Liddle and Cryer (1991) which comprise the fusion of mouse myeloma cells to spleen cells derived from immunized animals. Furthermore, antibodies or fragments thereof to a molecule or fragments thereof can be obtained by using methods as described in e.g. Harlow and Lane (1988). In the case of antibodies directed against small peptides such as fragments of a protein of the invention, said peptides are generally coupled to a carrier protein before immunization of animals. Such protein carriers include keyhole limpet hemocyanin (KLH), bovine serum albumin (BSA), ovalbumin and Tetanus toxoid. The carrier protein enhances the immune response of the animal and provides epitopes for T-cell receptor binding sites. The term "antibodies" furthermore includes derivatives thereof such as labelled antibodies. Antibody labels include alkaline phosphatase, PKH2, PKH26, PKH67, fluorescein (FITC), Hoecst 33258, R-phycoerythrin (PE), rhodamine (TRITC), Quantum Red, Texas Red, Cy3, biotin, agarose, peroxidase, gold

spheres and radiolabels (e.g. ^{125}I, ^{131}I, ^{35}S, ^{14}C, ^{32}P, ^{33}P, ^{3}H). Tools in molecular biology relying on antibodies against a protein include protein gel blot analysis, screening of expression libraries allowing gene identification, protein quantitative methods including ELISA and RIA, immunoaffinity purification of proteins, immunoprecipitation of proteins (e.g. Magyar et al. 1997) and immunolocalization. Other uses of antibodies and especially of peptide antibodies include the study of proteolytic processing (Loffler et al. 1994, Woulfe et al. 1994), determination of protein active sites (Lerner 1982), the study of precursor and post-translational processing (Baron and Baltimore 1982, Lerner et al. 1981, Semler et al. 1982), identification of protein domains involved in protein-protein interactions (Murakami et al. 1992) and the study of exon usage in gene expression (Tamura et al. 1991).

Embodied in the current invention are antibodies recognizing a LRR receptor-like kinase polypeptide or protein, or homolog, derivative or fragment thereof as defined supra or specific epitopes of said polypeptide or protein.

The terms "gene(s)", "polynucleotide(s)", "nucleic acid sequence(s)", "nucleotide sequence(s)", "DNA sequence(s)" or "nucleic acid molecule(s)", when used herein refer to nucleotides, either ribonucleotides or deoxyribonucleotides or a combination of both, in a polymeric form of any length. Said terms furthermore include double-stranded and single-stranded DNA and RNA. Said terms also include known nucleotide modifications such as methylation, cyclization and 'caps' and substitution of one or more of the naturally occurring nucleotides with an analog such as inosine. Modifications of nucleotides include the addition of acridine, amine, biotin, cascade blue, cholesterol, Cy3®, Cy5®, Cy5.5® Dabcyl, digoxigenin, dinitrophenyl, Edans, 6-FAM, fluorescein, 3'-glyceryl, HEX, IRD-700, IRD-800, JOE, phosphate psoralen, rhodamine, ROX, thiol (SH), spacers, TAMRA, TET, AMCA-S®, SE, BODIPY®, Marina Blue®, Pacific Blue®, Oregon Green®, Rhodamine Green®, Rhodamine Red®, Rhodol Green® and Texas Red®. Polynucleotide backbone modifications include methylphosphonate, 2'-OMe-methylphosphonate RNA, phosphorothiorate, RNA, 2'-OMeRNA. Base modifications include 2-amino-dA, 2-aminopurine, 3'-(ddA), 3'dA(cordycepin), 7-deaza-dA, 8-Br-dA, 8-oxo-dA, N^6-Me-dA, abasic site (dSpacer), biotin dT, 2'-OMe-5Me-C, 2'-OMe-propynyl-C, 3'-(5-Me-dC), 3'-(ddC), 5-Br-dC, 5-I-dC, 5-Me-dC, 5-F-dC, carboxy-dT, convertible dA, convertible dC, convertible dG, convertible dT, convertible dU, 7-deaza-dG, 8-Br-dG, 8-oxo-dG, O^6-Me-dG, S6-DNP-dG, 4-methyl-indole, 5-nitroindole, 2'-OMe-inosine, 2'-dI, 0^6-

phenyl-dI, 4-methyl-indole, 2'-deoxynebularine, 5-nitroindole, 2-aminopurine, dP(purine analogue), dK(pyrimidine analogue), 3-nitropyrrole, 2-thio-dT, 4-thio-dT, biotin-dT, carboxy-dT, O^4-Me-dT, O^4-triazol dT, 2'-OMe-propynyl-U, 5-Br-dU, 2'-dU, 5-F-dU, 5-I-dU, O^4-triazol dU and radiolabels (e.g. ^{125}I, ^{131}I, ^{35}S, ^{14}C, ^{32}P, ^{33}P, ^{3}H). Said terms also encompass peptide nucleic acids (PNAs), a DNA analogue in which the backbone is a pseudopeptide consisting of N-(2-aminoethyl)-glycine units rather than a sugar. PNAs mimic the behaviour of DNA and bind complementary nucleic acid strands. The neutral backbone of PNA results in stronger binding and greater specificity than normally achieved. In addition, the unique chemical, physical and biological properties of PNA have been exploited to produce powerful biomolecular tools, antisense and antigene agents, molecular probes and biosensors. With "recombinant DNA molecule" or "chimeric gene" is meant a hybrid DNA produced by joining pieces of DNA from different sources.With "heterologous nucleotide sequence" is intended a sequence that is not naturally occurring with the promoter sequence. While this nucleotide sequence is heterologous to the promoter sequence, it may be homologous, or native, or heterologous, or foreign, to the plant host. "Sense strand" refers to the strand of a double-stranded DNA molecule that is homologous to a mRNA transcript thereof. The "anti-sense strand" contains an inverted sequence which is complementary to that of the "sense strand".

A "coding sequence" or "open reading frame" or "ORF" is defined as a nucleotide sequence that can be transcribed into mRNA and/or translated into a polypeptide when placed under the control of appropriate regulatory sequences, i.e. when said coding sequence or ORF is present in an expressible format. Said coding sequence of ORF is bounded by a 5' translation start codon and a 3' translation stop codon. A coding sequence or ORF can include, but is not limited to RNA, mRNA, cDNA, recombinant nucleotide sequences, synthetically manufactured nucleotide sequences or genomic DNA. Said coding sequence or ORF can be interrupted by intervening nucleic acid sequences.

Table 2. Degeneracy of the genetic code.

Amino Acid	Three-letter code	One-letter code	Possible codons					
Alanine	Ala	A	GCA	GCC	GCG	GCU		
Arginine	Arg	R	AGA	AGG	CGA	CGC	CGG	CGU
Asparagine	Asn	N	AAC	AAU				
Aspartic Acid	Asp	D	GAC	GAU				
Cysteine	Cys	C	UGC	UGU				
Glutamic Acid	Glu	E	GAA	GAG				
Glutamine	Gln	Q	CAA	CAG				
Glycine	Gly	G	GGA	GGC	GGG	GGU		
Histidine	His	H	CAC	CAU				
Isoleucine	Ile	I	AUA	AUC	AUU			
Leucine	Leu	L	UUA	UUG	CUA	CUC	CUG	CUU
Lysine	Lys	K	AAA	AAG				
Methionine	Met	M	AUG					
Phenylalanine	Phe	F	UUC	UUU				
Proline	Pro	P	CCA	CCC	CCG	CCU		
Serine	Ser	S	AGC	AGU	UCA	UCC	UCG	UCU
Threonine	Thr	T	ACA	ACC	ACG	ACU		
Tryptophan	Trp	W	UGG					
Tyrosine	Tyr	Y	UAC	UAU				
Valine	Val	V	GUA	GUC	GUG	GUU		
			Possible "STOP" codons					
			UAA	UAG	UGA			

Genes and coding sequences essentially encoding the same protein but isolated from different sources can consist of substantially divergent nucleic acid sequences. Reciprocally, substantially divergent nucleic acid sequences can be designed to effect expression of essentially the same protein. Said nucleic acid sequences are the result of e.g. the existence of different alleles of a given gene, of the degeneracy of the genetic code or of differences in codon usage. Thus, as indicated in Table 2, amino acids such

as methionine and tryptophan are encoded by a single codon whereas other amino acids such as arginine, leucine and serine can each be translated from up to six different codons. Differences in preferred codon usage are illustrated below for *Agrobacterium tumefaciens* (a bacterium), *A. thaliana*, *M. sativa* (two dicotyledonous plants) and *Oryza sativa* (a monocotyledonous plant). These examples were extracted from (http://www.kazusa.or.jp/codon). To give one example, the codon GGC (for glycine) is the most frequently used codon in *A. tumefaciens* (36.2 ‰), is the second most frequently used codon in *O. sativa* but is used at much lower frequencies in *A. thaliana* and *M. sativa* (9 ‰ and 8.4 ‰, respectively). Of the four possible codons encoding glycine (see Table 6), said GGC codon is most preferably used in *A. tumefaciens* and *O. sativa*. However, in *A. thaliana* this is the GGA (and GGU) codon whereas in *M. sativa* this is the GGU (and GGA) codon.

"Hybridization" is the process wherein substantially homologous complementary nucleotide sequences anneal to each other. The hybridization process can occur entirely in solution, i.e. both complementary nucleic acids are in solution. Tools in molecular biology relying on such a process include the polymerase chain reaction (PCR; and all methods based thereon), subtractive hybridization, random primer extension, nuclease S1 mapping, primer extension, reverse transcription, cDNA synthesis, differential display of RNAs, and DNA sequence determination. The hybridization process can also occur with one of the complementary nucleic acids immobilized to a matrix such as magnetic beads, Sepharose beads or any other resin. Tools in molecular biology relying on such a process include the isolation of poly (A+) mRNA. The hybridization process can furthermore occur with one of the complementary nucleic acids immobilized to a solid support such as a nitrocellulose or nylon membrane or immobilized by e.g. photolitography to e.g. a silicious glass support (the latter known as nucleic acid arrays or microarrays or as nucleic acid chips). Tools in molecular biology relying on such a process include RNA and DNA gel blot analysis, colony hybridization, plaque hybridization, in situ hybridization and microarray hybridization. In order to allow hybridization to occur, the nucleic acid molecules are generally thermally or chemically denatured to melt a double strand into two single strands and/or to remove hairpins or other secondary structures from single stranded nucleic acids. The stringency of hybridization is influenced by conditions such as temperature, salt concentration and hybridization buffer composition. High stringency conditions for hybridization include high

temperature and/or low salt concentration (salts include NaCl and Na_3-citrate) and/or the inclusion of formamide in the hybridization buffer and/or lowering the concentration of compounds such as SDS (detergent) in the hybridization buffer and/or exclusion of compounds such as dextran sulfate or polyethylene glycol (promoting molecular crowding) from the hybridization buffer. Conventional hybridization conditions are described in e.g. Sambrook et al. (1989) but the skilled craftsman will appreciate that numerous different hybridization conditions can be designed in function of the known or the expected homology and/or length of the nucleic acid sequence. Sufficiently low stringency hybridization conditions are particularly preferred to isolate nucleic acids heterologous to the DNA sequences of the invention defined supra. Elements contributing to said heterology include allelism, degeneration of the genetic code and differences in preferred codon usage as discussed supra.

Clearly, the current invention embodies the use of the inventive DNA sequences encoding a LRR receptor-like kinase, homolog, derivative and/or immunologically fragment thereof as defined higher in any method of hybridization . The current invention furthermore also relates to DNA sequences specifically hybridizing to said inventive DNA sequences.

DNA sequences as defined in the current invention can be interrupted by intervening sequences. With "intervening sequences" is meant any nucleic acid sequence which disrupts a coding sequence comprising said inventive DNA sequence or which disrupts the expressible format of a DNA sequence comprising said inventive DNA sequence. Removal of the intervening sequence restores said coding sequence or said expressible format. Examples of intervening sequences include introns, mobilizable DNA sequences such as transposons and DNA tags such as e.g. a T-DNA. With "mobilizable DNA sequence" is meant any DNA sequence that can be mobilized as the result of a recombination event.

To effect expression of a protein in a cell, tissue or organ, preferably of plant origin, either the protein may be introduced directly to said cell, such as by microinjection or ballistic means or alternatively, an isolated nucleic acid molecule encoding said protein may be introduced into said cell, tissue or organ in an expressible format.

Preferably, the DNA sequence of the invention comprises a coding sequence or open reading frame (ORF) encoding a LRR receptor-like kinase or a homolog or derivative thereof or an immunologically active fragment thereof as defined supra. The preferred

protein of the invention comprises the amino acid sequence of said LRR receptor-like kinase.

With "vector" or "vector sequence" is meant a DNA sequence which can be introduced in an organism by transformation and can be stably maintained in said organism. Vector maintenance is possible in e.g. cultures of *Escherichia coli*, *A. tumefaciens*, *Saccharomyces cerevisiae* or *Schizosaccharomyces pombe*. Other vectors such as phagemids and cosmid vectors can be maintained and multiplied in bacteria and/or viruses. Vector sequences generally comprise a set of unique sites recognized by restriction enzymes, the multiple cloning site (MCS), wherein one or more non-vector sequence(s) can be inserted.

With "non-vector sequence" is accordingly meant a DNA sequence which is integrated in one or more of the sites of the MCS comprised within a vector.

"Expression vectors" form a subset of vectors which, by virtue of comprising the appropriate regulatory sequences enabling the creation of an expressible format for the inserted non-vector sequence(s), thus allowing expression of the protein encoded by said non-vector sequence(s). Expression vectors are known in the art enabling protein expression in organisms including bacteria (e.g. *E. coli*), fungi (e.g. *S. cerevisiae*, *S. pombe*, *Pichia pastoris*), insect cells (e.g. baculoviral expression vectors), animal cells (e.g. COS or CHO cells) and plant cells (e.g. potato virus X-based expression vectors, see e.g. Vance et al. 1998 - WO9844097).

The current invention clearly includes any vector or expression vector comprising a non-vector DNA sequence comprising the promoter sequence according to the present invention or a non)vector sequence encoding a LRR receptor-like kinase, homolog, derivative and/or immunologically active fragment thereof as defined supra.

As an alternative to expression vector-mediated protein production in biological systems, chemical protein synthesis can be applied. Synthetic peptides can be manufactured in solution phase or in solid phase. Solid phase peptide synthesis (Merrifield 1963) is, however, the most common way and involves the sequential addition of amino acids to create a linear peptide chain. Solid phase peptide synthesis includes cycles consisting of three steps: (i) immobilization of the carboxy-terminal amino acid of the growing peptide chain to a solid support or resin; (ii) chain assembly, a process consisting of activation, coupling and deprotection of the amino acid to be added to the growing peptide chain; and (iii) cleavage involving removal of the completed peptide chain from the resin and

removal of the protecting groups from the amino acid side chains. Common approaches in solid phase peptide synthesis include Fmoc/tBu (9-fluorenylmethyloxycarbonyl/t-butyl) and Boc (t-butyloxycarbonyl) as the amino-terminal protecting groups of amino acids. Amino acid side chain protecting groups include methyl (Me), formyl (CHO), ethyl (Et), acetyl (Ac), t-butyl (t-Bu), anisyl, benzyl (Bzl), trifluroacetyl (Tfa), N-hydroxysuccinimide (ONSu, OSu), benzoyl (Bz), 4-methylbenzyl (Meb), thioanizyl, thiocresyl, benzyloxymethyl (Bom), 4-nitrophenyl (ONp), benzyloxycarbonyl (Z), 2-nitrobenzoyl (NBz), 2-nitrophenylsulphenyl (Nps), 4-toluenesulphonyl (Tosyl,Tos), pentafluorophenyl (Pfp), diphenylmethyl (Dpm), 2-chlorobenzyloxycarbonyl (Cl-Z), 2,4,5-trichlorophenyl, 2-bromobenzyloxycarbonyl (Br-Z), tripheylmethyl (Trityl, Trt), and 2,5,7,8-pentamethyl-chroman-6-sulphonyl (Pmc). During chain assembly, Fmoc or Boc are removed resulting in an activated amino-terminus of the amino acid residue bound to the growing chain. The carboxy-terminus of the incoming amino acid is activated by conversion into a highly reactive ester, e.g. by HBTU. With current technologies (e.g. PerSeptive Biosystems 9050 synthesizer, Applied Biosystems Model 431A Peptide Synthesizer), linear peptides of up to 50 residues can be manufactured. A number of guidelines is available to produce peptides that are suitable for use in biological systems including (i) limiting the use of difficult amino acids such as cys, met, trp (easily oxidized and/or degraded during peptide synthesis) or arg; (ii) minimize hydrophobic amino acids (can impair peptide solubility); and (iii) prevent an amino-terminal glutamic acid (can cyclize to pyroglutamate).

By "expressible format" is meant that the isolated nucleic acid molecule is in a form suitable for being transcribed into mRNA and/or translated to produce a protein, either constitutively or following induction by an intracellular or extracellular signal, such as an environmental stimulus or stress (mitogens, anoxia, hypoxia, temperature, salt, light, dehydration, etc) or a chemical compound such as IPTG (isopropyl-β-D-thiogalactopyranoside) or such as an antibiotic (tetracycline, ampicillin, rifampicin, kanamycin), hormone (e.g. gibberellin, auxin, cytokinin, glucocorticoid, brassinosteroid, ethylene, abscisic acid etc), hormone analogue (iodoacetic acid (IAA), 2,4-D, etc), metal (zinc, copper, iron, etc), or dexamethasone, amongst others. As will be known to those skilled in the art, expression of a functional protein may also require one or more post-translational modifications, such as glycosylation, phosphorylation, dephosphorylation,

or one or more protein-protein interactions, amongst others. All such processes are included within the scope of the term "expressible format".

Preferably, expression of a protein in a specific cell, tissue, or organ, preferably of plant origin, is effected by introducing and expressing an isolated nucleic acid molecule encoding said protein, such as a cDNA molecule, genomic gene, synthetic oligonucleotide molecule, mRNA molecule or open reading frame, to said cell, tissue or organ, wherein said nucleic acid molecule is placed operably in connection with suitable regulatory sequences including a promoter, preferably a plant-expressible promoter, and a terminator sequence.

"Regulatory sequence" refers to control DNA sequences which are necessary to affect the expression of coding sequences to which they are ligated. The nature of such control sequences differs depending upon the host organism. In prokaryotes, control sequences generally include promoters, ribosomal binding sites, and terminators. In eukaryotes generally control sequences include promoters, terminators and enhancers or silencers. The term "control sequence" is intended to include, at a minimum, all components the presence of which are necessary for expression, and may also include additional advantageous components and which determines when, how much and where a specific gene is expressed.

Reference herein to a "promoter" is to be taken in its broadest context and includes the transcriptional regulatory sequences derived from a classical eukaryotic genomic gene, including the TATA box which is required for accurate transcription initiation, with or without a CCAAT box sequence and additional regulatory elements (i.e. upstream activating sequences, enhancers and silencers) which alter gene expression in response to developmental and/or external stimuli, or in a tissue-specific manner. The term "promoter" also includes the transcriptional regulatory sequences of a classical prokaryotic gene, in which case it may include a -35 box sequence and/or a -10 box transcriptional regulatory sequences.

The term "promoter" is also used to describe a synthetic or fusion molecule, or derivative which confers, activates or enhances expression of a nucleic acid molecule in a cell, tissue or organ.

The term "functional part" when referred to nucleic acids in relation to the above defined transcriptional regulatory sequences, refers to a part or parts of the nucleic acid having

the activity to specifically drive or promote transcription from sequences which are located downstream of said nucleic acid sequence.

Promoters may contain additional copies of one or more specific regulatory elements, to further enhance expression and/or to alter the spatial expression and/or temporal expression of a nucleic acid molecule to which it is operably connected. Such regulatory elements may be placed adjacent to a heterologous promoter sequence to drive expression of a nucleic acid molecule in response to e.g. copper, glucocorticoids, dexamethasone, tetracycline, gibberellin, cAMP, abscisic acid, auxin, wounding, ethylene, jasmonate or salicylic acid or to confer expression of a nucleic acid molecule to specific cells, tissues or organs such as meristems, leaves, roots, embryo, flowers, seeds or fruits.

In the context of the present invention, the promoter preferably is a plant-expressible promoter sequence. Promoters, however, that also function or solely function in non-plant cells such as bacteria, yeast cells, insect cells and animal cells are not excluded from the invention. By "plant-expressible" is meant that the promoter sequence, including any additional regulatory elements added thereto or contained therein, is at least capable of inducing, conferring, activating or enhancing expression in a plant cell, tissue or organ, preferably a monocotyledonous or dicotyledonous plant cell, tissue, or organ. The terms "plant-operable" and "operable in a plant" when used herein, in respect of a promoter sequence, shall be taken to be equivalent to a plant-expressible promoter sequence.

In the present context, a "regulatable promoter sequence" is a promoter that is capable of conferring expression on a structural gene in a particular cell, tissue, or organ or group of cells, tissues or organs of a plant, optionally under specific conditions, however does generally not confer expression throughout the plant under all conditions. Accordingly, a regulatable promoter sequence may be a promoter sequence that confers expression on a gene to which it is operably connected in a particular location within the plant or alternatively, throughout the plant under a specific set of conditions, such as following induction of gene expression by a chemical compound or other elicitor.

Preferably, the regulatable promoter used in the performance of the present invention confers expression in a specific location within the plant, either constitutively or following induction, however not in the whole plant under any circumstances. Included within the scope of such promoters are cell-specific promoter sequences, tissue-specific promoter

sequences, organ-specific promoter sequences, cell cycle specific gene promoter sequences, inducible promoter sequences and constitutive promoter sequences that have been modified to confer expression in a particular part of the plant at any one time, such as by integration of said constitutive promoter within a transposable genetic element (*Ac, Ds, Spm, En*, or other transposon). The skilled craftsman will understand that a "constitutive promoter" is a promoter that is transcriptionaly active in an organism, preferably a plant, during most, but not necessarily all phases of its growth and development. Similarly, the skilled craftsman will understand that a "ubiquitous promoter" is a promoter that is transcriptionally active throughout most, but not necessarily all parts of an organism, preferably a plant.

Generally by "weak promoter" is intended a promoter that drives expression of a coding sequence at a low level. By "low level" is intended at levels of about 1/10,000 transcripts to about 1/100,000 transcripts, to about 1/500,0000 transcripts. Conversely, a "strong promoter" drives expression of a coding sequence at high level, or at about 1/10 transcripts to about 1/100 transcripts to about 1/1,000 transcripts.

The term "cell-specific" shall be taken to indicate that expression is predominantly in a particular cell or cell-type, preferably of plant origin, albeit not necessarily exclusively in said cell or cell-type.

Similarly, the term "tissue-specific" shall be taken to indicate that expression is predominantly in a particular tissue or tissue-type, preferably of plant origin, albeit not necessarily exclusively in said tissue or tissue-type.

Similarly, the term "organ-specific" shall be taken to indicate that expression is predominantly in a particular organ, preferably of plant origin, albeit not necessarily exclusively in said organ. "Root specific" means that the promoter is expressed in the root only and not in other tissues of the plant.

By "root-preferred" it is intended that expression of the heterologous nucleotide sequence is most abundant root, but could also have low expression levels elsewhere in the plant. While some level of expression of the heterologous nucleotide sequence occurs in other plant tissue types, expression occurs most abundantly in the root including primary, lateral and adventitious roots.

By "root" is intended any part of the root structure, including, but not limited to, the root cap, apical meristem, protoderm, ground meristem, procambium, endodermis, cortex, vascular cortex, epidermis, and the like.

Similarly, the term "cell cycle specific" shall be taken to indicate that expression is predominantly cyclic and occurring in one or more, not necessarily consecutive phases of the cell cycle albeit not necessarily exclusively in cycling cells, preferably of plant origin.

Contrarily, the term ubiquitous shall be taken to indicate that expression is throughout the entire oranism.

Those skilled in the art will be aware that an "inducible promoter" is a promoter the transcriptional activity of which is increased or induced in response to a developmental, chemical, environmental, or physical stimulus. Similarly, the skilled craftsman will understand that a "constitutive promoter" is a promoter that is transcriptionally active throughout most, but not necessarily all parts of an organism, preferably a plant, during most, but not necessarily all phases of its growth and development.

Those skilled in the art will readily be capable of selecting appropriate promoter sequences for use in regulating appropriate expression of the LRR receptor-like kinase as described supra from publicly-available or readily-available sources, without undue experimentation.

Placing a nucleic acid molecule under the regulatory control of a promoter sequence, or in operable connection with a promoter sequence, means positioning said nucleic acid molecule such that expression is controlled by the promoter sequence. A promoter is usually, but not necessarily, positioned upstream, or at the 5'-end, and within 2 kb of the start site of transcription, of the nucleic acid molecule which it regulates. In the construction of heterologous promoter/structural gene combinations it is generally preferred to position the promoter at a distance from the gene transcription start site that is approximately the same as the distance between that promoter and the gene it controls in its natural setting (i.e., the gene from which the promoter is derived). As is known in the art, some variation in this distance can be accommodated without loss of promoter function. Similarly, the preferred positioning of a regulatory sequence element with respect to a heterologous gene to be placed under its control is defined by the positioning of the element in its natural setting (i.e., the gene from which it is derived). Again, as is known in the art, some variation in this distance can also occur.

"Expression" means the production of a protein or nucleotide sequence in the cell itself or in a cell-free system. It includes transcription into an RNA product, post-transcriptional

modification and/or translation to a protein product or polypeptide from a DNA encoding that product, as well as possible post-translational modifications.

"Operably linked" refers to a juxtaposition wherein the components so described are in a relationship permitting them to function in their intended manner. A control sequence "operably linked" to a coding sequence is ligated in such a way that expression of the coding sequence is achieved under conditions compatible with the control sequences. In case the control sequence is a promoter, it is obvious for a skilled person that double-stranded nucleic acid is preferably used.

Examples of promoters suitable for use in gene constructs of the present invention include those listed in Table 3, amongst others. The promoters listed in Table 5 are provided for the purposes of exemplification only and the present invention is not to be limited by the list provided therein. Those skilled in the art will readily be in a position to provide additional promoters that are useful in performing the present invention.

In the case of constitutive promoters or promoters that induce expression throughout the entire plant, it is preferred that such sequences are modified by the addition of nucleotide sequences derived from one or more of the tissue-specific promoters listed in Table 8, or alternatively, nucleotide sequences derived from one or more of the above-mentioned tissue-specific inducible promoters, to confer tissue-specificity thereon. For example, the CaMV 35S promoter may be modified by the addition of maize *Adh1* promoter sequence, to confer anaerobically-regulated root-specific expression thereon, as described previously (Ellis *et al.*, 1987). Another example describes conferring root specific or root abundant gene expression by fusing the CaMV35S promoter to elements of the maize glycine-rich protein GRP3 gene (Feix and Wulff 2000 - WO0015662). Such modifications can be achieved by routine experimentation by those skilled in the art.

The term "terminator" refers to a DNA sequence at the end of a transcriptional unit which signals termination of transcription. Terminators are 3'-non-translated DNA sequences containing a polyadenylation signal, which facilitates the addition of polyadenylate sequences to the 3'-end of a primary transcript. Terminators active in cells derived from viruses, yeasts, moulds, bacteria, insects, birds, mammals and plants are known and described in the literature. They may be isolated from bacteria, fungi, viruses, animals and/or plants.

Examples of terminators particularly suitable for use in the gene constructs of the present invention include the *Agrobacterium tumefaciens* nopaline synthase (NOS) gene

terminator, the *Agrobacterium tumefaciens* octopine synthase (OCS) gene terminator sequence, the Cauliflower mosaic virus (CaMV) 35S gene terminator sequence, the *Oryza sativa* ADP-glucose pyrophosphorylase terminator sequence (t3'Bt2), the *Zea mays* zein gene terminator sequence, the *rbcs-1A* gene terminator, and the *rbcs-3A* gene terminator sequences, amongst others.

Table 3. Exemplary plant-expressible promoters for use in the performance of the present invention

I: CELL-SPECIFIC,TISSUE-SPECIFIC, AND ORGAN-SPECIFIC PROMOTERS		
GENE SOURCE	EXPRESSION PATTERN	REFERENCE
α-amylase (*Amy32b*)	Aleurone	Lanahan et al, Plant Cell 4:203-211, 1992; Skriver et al, Proc Natl Acad Sci USA 88:7266-7270, 1991
cathepsin β-like gene	Aleurone	Cejudo et al, Plant Mol Biol 20:849-856, 1992
Agrobacterium rhizogenes rolB	Cambium	Nilsson et al, Physiol Plant 100:456-462, 1997
AtPRP4	Flowers	http://salus.medium.edu/mmg/tierney/html
chalcone synthase (chsA)	Flowers	Van der Meer et al, Plant Mol Biol 15:95-109, 1990
LAT52	Anther	Twell et al, Mol Gen Genet 217:240-245, 1989
apetala-3	Flowers	
chitinase	fruit (berries, grapes, etc)	Thomas et al. CSIRO Plant Industry, Urrbrae, South Australia, Australia; http://winetitles.com.au/gwrdc/csh95-1.html
rbcs-3A	green tissue (eg leaf)	Lam et al, Plant Cell 2:857-866, 1990; Tucker et al., Plant Physiol 113:1303-1308, 1992
leaf-specific genes	Leaf	Baszczynski et al, Nucl Acid Res 16:4732, 1988
AtPRP4	Leaf	http://salus.medium.edu/mmg/tierney/html
chlorella virus adenine methyltransferase gene promoter	Leaf	Mitra and Higgins, Plant Mol Biol 26:85-93, 1994
aldP gene promoter from rice	Leaf	Kagaya et al, Mol Gen Genet 248:668-674, 1995

rbcs promoter from rice or tomato	Leaf	Kyozuka et al, Plant Physiol 102:991-1000, 1993
Pinus cab-6	Leaf	Yamamoto et al, Plant Cell Physiol 35:773-778, 1994
rubisco promoter	Leaf	
cab (chlorophyll a/b/binding protein	Leaf	
pea Blec4 gene	Vegetative and floral epidermal tissues	Mandaci and Dobres, Plant Mol Biol 34:961-965
SAM22	senescent leaf	Crowell et al, Plant Mol Biol 18:459-466, 1992
ltp gene (lipid transfer gene)		Fleming et al, Plant J 2:855-862, 1992
R. japonicum nif gene	nodule	United States Patent No 4 803165
B. japonicum nifH gene	nodule	United States Patent No 5008194
GmENOD40	nodule	Yang et al, Plant J 3:573-585, 1993
PEP carboxylase (PEPC)	nodule	Pathirana et al, Plant Mol Biol 20:437-450, 1992
Leghaemoglobin (Lb)	nodule	Gordon et al, J Exp Bot 44:1453-1465, 1993
Tungro bacilliform virus gene	phloem	Bhattacharyya-Pakrasi et al, Plant J 4:71-79, 1992
pollen-specific genes	pollen; microspore	Albani et al, Plant Mol Biol 15:605, 1990; Albani et al, Plant Mol Biol 16:501, 1991
Zm13	pollen	Guerrero et al, Mol Gen Genet 224:161-168, 1993
apg gene	microspore	Twell et al, Sex Plant Reprod 6:217-224, 1993
maize pollen-specific gene	pollen	Hamilton et al, Plant Mol Biol 18:211-218, 1992
Sunflower pollen-expressed gene	pollen	Baltz et al, Plant J 2:713-721, 1992
B. napus pollen-	pollen;anther; tapetum	Arnoldo et al, J Cell Biochem, Abstract No.

specific gene		Y101, 204, 1992
root-expressible genes	roots	Tingey et al, EMBO J 6:1, 1987
Tobacco auxin-inducible gene	root tip	Van der Zaal et al, Plant Mol Biol 16:983, 1991
β-tubulin	root	Oppenheimer et al, Gene 63:87, 1988
Tobacco root-specific genes	root	Conkling et al, Plant Physiol 93:1203, 1990
B. napus G1-3b gene	root	United States Patent No 5401836
SbPRP1	roots	Suzuki et al, Plant Mol Biol 21:109-119, 1993
AtPRP1; AtPRP3	roots; root hairs	http://salus.medium.edu/mmg/tierney/html
RD2 gene	root cortex	http://www2.cnsu.edu/ncsu/research
TobRB7 gene	root vasculature	http://www2.cnsu.edu/ncsu/research
AtPRP4	leaves; flowers; lateral root primordia	http://salus.medium.edu/mmg/tierney/html
seed-specific genes	seed	Simon et al, Plant Mol Biol 5:191, 1985; Scofield et al, J Biol Chem 262:12202, 1987; Baszczynski et al, Plant Mol Biol 14:633, 1990
Brazil Nut albumin	seed	Pearson et al, Plant Mol Biol 18:235-245, 1992
Legumin	seed	Ellis et al, Plant Mol Biol 10:203-214, 1988
Glutelin (rice)	seed	Takaiwa et al, Mol Gen Genet 208:15-22, 1986; Takaiwa et al, FEBS Lett 221:43-47, 1987
Zein	seed	Matzke et al, Plant Mol Biol 14:323-32 1990
NapA	seed	Stalberg et al, Planta 199:515-519, 1996
wheat LMW and HMW glutenin-1	endosperm	Mol Gen Genet 216:81-90, 1989; Nucl Acids Res 17:461-462, 1989
wheat SPA	seed	Albani et al, Plant Cell 9:171-184, 1997
cZ19B1, maize 19 kDa zein	seed	WO0011177
mi1ps, maize	seed	WO0011177

myoinositol-1-PI synthase		
wheat α, β, γ-gliadins	endosperm	EMBO J 3:1409-1415, 1984
barley *ltr1* promoter	endosperm	
barley B1, C, D, hordein	endosperm	Theor Appl Gen 98:1253-1262, 1999; Plant J 4:343-355, 1993; Mol Gen Genet 250:750-60, 1996
barley DOF	endosperm	Mena et al, Plant J 116:53-62, 1998
blz2	endosperm	EP99106056.7
synthetic promoter	endosperm	Vicente-Carbajosa et al, Plant J 13:629-640, 1998
rice prolamin NRP33	endosperm	Wu et al, Plant Cell Physiol 39: 885-889, 1998
rice α-globulin Glb-1	endosperm	Wu et al, Plant Cell Physiol 39:885-889, 1998
maize END genes	endosperm	WO0012733
barley END1	endosperm	WO9808961
barley NUC1	nucellus	WO9808961
rice OSH1	embryo	Sato et al, Proc Natl Acad Sci USA 93:8117-8122, 1996
rice α-globulin REB/OHP-1	endosperm	Nakase et al, Plant Mol Biol 33:513-522, 1997
rice ADP-glucose PP	endosperm	Trans Res 6:157-168, 1997
maize ESR gene family	endosperm	Plant J 12:235-246, 1997
sorgum γ-kafirin	endosperm	Plant Mol Biol 32:1029-1035, 1996
KNOX	embryo	Postma-Haarsma et al, Plant Mol Biol 39:257-271, 1999
rice oleosin	embryo and aleuron	Wu et al, J Biochem 123:386, 1998
sunflower oleosin	seed (embryo and dry seed)	Cummins et al, Plant Mol Biol 19:873-876, 1992
LEAFY	shoot meristem	Weigel et al, Cell 69:843-859, 1992
Arabidopsis thaliana knat1	shoot meristem	Accession number AJ131822
Malus domestica	shoot meristem	Accession number Z71981

kn1		
CLAVATA1	shoot meristem	Accession number AF049870
stigma-specific genes	stigma	Nasrallah et al, Proc Natl Acad Sci USA 85:5551, 1988; Trick et al, Plant Mol Biol 15:203, 1990
class I patatin gene	tuber	Liu et al, Plant Mol Biol 153:386-395, 1991
PCNA rice	meristem	Kosugi et al, Nucl Acids Res 19:1571-1576, 1991; Kosugi and Ohashi, Plant Cell 9:1607-1619, 1997
Pea TubA1 tubulin	Dividing cells	Stotz and Long, Plant Mol Biol 41:601-614, 1999
Arabidopsis cdc2a	cycling cells	Chung and Parish, FEBS Lett 362:215-219, 1995
Arabidopsis Rop1A	Anthers; mature pollen + pollen tubes	Li et al, Plant Physiol 118:407-417, 1998
Arabidopsis AtDMC1	Meiosis-associated	Klimyuk and Jones, Plant J 11:1-14, 1997
Pea PS-IAA4/5 and PS-IAA6	Auxin-inducible	Wong et al, Plant J 9:587-599, 1996
Pea farnesyltransferase	Meristematic tissues; phloem near growing tissues; light- and sugar-repressed	Zhou et al, Plant J 12:921-930,1997
Tobacco (*N. sylvestris*) cyclin B1;1	Dividing cells / meristematic tissue	Trehin et al, Plant Mol.Biol. 35:667-672, 1997
Catharanthus roseus Mitotic cyclins CYS (A-type) and CYM (B-type)	Dividing cells / meristematic tissue	Ito et al, Plant J 11:983-992, 1997
Arabidopsis cyc1At (=cyc B1;1) and cyc3aAt (A-type)	Dividing cells / meristematic tissue	Shaul et al, Proc Natl Acad Sci USA 93:4868-4872, 1996
Arabidopsis tef1 promoter box	Dividing cells / meristematic tissue	Regad et al, Mol Gen Genet 248:703-711, 1995
Catharanthus roseus	Dividing cells /	Ito et al, Plant Mol Biol 24:863-878, 1994

cyc07	meristematic tissue	
II: EXEMPLARY CONSTITUTIVE PROMOTERS		
GENE SOURCE	**EXPRESSION PATTERN**	**REFERENCE**
Actin	constitutive	McElroy et al, Plant Cell 2:163-171, 1990
CAMV 35S	constitutive	Odell et al, Nature 313:810-812, 1985
CaMV 19S	constitutive	Nilsson et al, Physiol Plant 100:456-462, 1997
GOS2	constitutive	de Pater et al, Plant J 2:837-844, 1992
Ubiquitin	constitutive	Christensen et al, Plant Mol Biol 18:675-689, 1992
rice cyclophilin	constitutive	Buchholz et al, Plant Mol Biol 25:837-843, 1994
maize histone H3	constitutive	Lepetit et al, Mol Gen Genet 231:276-285, 1992
alfalfa histone H3	constitutive	Wu et al, Nucleic Acids Res 17:3057-3063, 1989; Wu et al, Plant Mol Biol 11:641-649, 1988
actin 2	constitutive	An et al, Plant J 10:107-121, 1996
III: EXEMPLARY STRESS-INDUCIBLE PROMOTERS		
NAME	**STRESS**	**REFERENCE**
P5CS (delta(1)-pyrroline-5-carboxylate syntase)	salt, water	Zhang et al, Plant Sci 129:81-89, 1997
cor15a	cold	Hajela et al, Plant Physiol 93:1246-1252, 1990
cor15b	cold	Wilhelm et al, Plant Mol Biol 23:1073-1077, 1993
cor15a (-305 to +78 nt)	cold, drought	Baker et al, Plant Mol Biol 24: 01-713, 1994
rd29	salt, drought, cold	Kasuga et al, Nature Biotechnol 18:287-291, 1999
heat shock proteins,	heat	Barros et al, Plant Mol Biol 19 665-75,

including artificial promoters containing the heat shock element (HSE)		1992. Marrs et al, Dev Genet14:27-41, 1993. Schoffl et al, Mol Gen Genet 217:246-53, 1989.
SmHSP (small heat shock proteins)	heat	Waters et al, J Exp Bot 47:325-338, 1996
wcs120	cold	Ouellete et al, FEBS Lett 423:324-328, 1998
ci7	cold	Kirch et al, Plant Mol Biol 33:897-909, 1997
Adh	cold, drought, hypoxia	Dolferus et al, Plant Physiol 105:1075-87, 1994
pwsi18	salt and drought	Joshee et al, Plant Cell Physiol 39:64-72, 1998
ci21A	cold	Schneider et al, Plant Physiol 113:335-45, 1997
Trg-31	drought	Chaudhary et al, Plant Mol Biol 30:1247-57, 1996
Osmotin	osmotic	Raghothama et al, Plant Mol Biol 23:1117-28, 1993
LapA	wounding, enviromental	WO99/03977 University of California/INRA
IV: EXEMPLARY PATHOGEN-INDUCIBLE PROMOTERS		
NAME	**PATHOGEN**	**REFERENCE**
RB7	Root-knot nematodes (Meloidogyne spp.)	US5760386 - North Carolina State University; Opperman et al, Science 263:221-23, 1994
PR-1, 2, 3, 4, 5, 8, 11	fungal, viral, bacterial	Ward et al, Plant Cell 3:1085-1094, 1991; Reiss et al 1996; Lebel et al, Plant J 16:223-233, 1998; Melchers et al, Plant J 5:469-480, 1994; Lawton et al, Plant Mol Biol, 19:735-743, 1992
HMG2	nematodes	WO9503690 - Virginia Tech Intellectual Properties Inc .
Abi3	Cyst nematodes	unpublished

	(Heterodera spp.)	
ARM1	nematodes	Barthels et al, Plant Cell 9:2119-2134, 1997 WO 98/31822 – Plant Genetic Systems
Att0728	nematodes	Barthels et al, Plant Cell 9: 2119-2134, 1997 *PCT/EP98/07761*
Att1712	nematodes	Barthels et al, Plant Cell 9, 2119-2134, 1997 *PCT/EP98/07761*
Gst1	Different types of pathogens	Strittmatter et al, Mol Plant-Microbe Interact 9:68-73, 1996
LEMMI	nematodes	*WO 92/21757 – Plant Genetic Systems*
CLE	geminivirus	*PCT/EP99/03445 - CINESTAV*
PDF1.2	Fungal including *Alternaria brassicicola* and *Botrytis cinerea*	Manners et al, Plant Mol Biol, 38:1071-1080, 1998
Thi2.1	Fungal – *Fusarium oxysporum f sp. matthiolae*	Vignutelli et al, Plant J 14:285-295, 1998
DB#226	nematodes	Bird and Wilson, Mol Plant-Microbe Interact 7:419-442, 1994 WO 95.322888
DB#280	nematodes	Bird and Wilson, Mol Plant-Microbe Interact 7:419-442, 1994 WO 95.322888
Cat2	nematodes	Niebel et al, Mol Plant-Microbe Interact 8:371-378, 1995
□Tub	nematodes	Aristizabal et al (1996), 8th International Congress on Plant-Microbe Interaction, Knoxville US B-29
SHSP	nematodes	Fenoll et al (1997) In: Cellular and molecular aspects of plant-nematode interactions. Kluwer Academic, C. Fenoll, F.M.W. Grundler and S.A. Ohl (Eds.),
Tsw12	nematodes	Fenoll et al (1997) In: Cellular and

		molecular aspects of plant-nematode interactions. Kluwer Academic, C. Fenoll, F.M.W. Grundler and S.A. Ohl (Eds.)
Hs1(pro1)	nematodes	WO 98/122335 - Jung
NsLTP	viral, fungal, bacterial	Molina and Garcia-Olmedo FEBS Lett, 316:119-122, 1993
RIP	viral, fungal	Tumer et al, Proc Natl Acad Sci USA 94:3866-3871, 1997

Those skilled in the art will be aware of additional promoter sequences and terminator sequences which may be suitable for use in performing the invention. Such sequences may readily be used without any undue experimentation.

In the context of the current invention, "ectopic expression" or "ectopic overexpression" of a gene or a protein are conferring to expression patterns and/or expression levels of said gene or protein normally not occurring under natural conditions. Ectopic expression can be achieved in a number of ways including operably linking of a coding sequence encoding said protein to an isolated homologous or heterologous promoter in order to create a chimeric gene and/or operably linking said coding sequence to its own isolated promoter (i.e. the unisolated promoter naturally driving expression of said protein) in order to create a recombinant gene duplication or gene multiplication effect. With "ectopic co-expression" is meant the ectopic expression or ectopic overexpression of two or more genes or proteins. The same or, more preferably, different promoters are used to confer expression of said genes or proteins.

Preferably, the promoter sequence used in the context of the present invention is operably linked to a coding sequence or open reading frame (ORF) encoding a LRR receptor-like kinase or a homolog, derivative and/or an immunologically active fragment thereof as defined supra.

"Dominant negative version or variant" refers to a mutant protein, which interferes with the activity of the corresponding wild-type protein.

"Downregulation of expression" as used herein means lowering levels of gene expression and/or levels of active gene product and/or levels of gene product activity. Decreases in expression may be accomplished by e.g. the addition of coding sequences or parts thereof in a sense orientation (if resulting in co-suppression) or in an antisense

orientation relative to a promoter sequence and furthermore by e.g. insertion mutagenesis (e.g. T-DNA insertion or transposon insertion) or by gene silencing strategies as described by e.g. Angeli and Baulcombe (1998 - WO9836083), Lowe et al. (1989 - WO9853083), Lederer et al. (1999 - WO9915682) or Wang et al. (1999 - WO9953050). Genetic constructs aimed at silencing gene expression may have the nucleotide sequence of said gene (or one or more parts thereof) contained therein in a sense and/or antisense orientation relative to the promoter sequence. Another method to downregulate gene expression comprises the use of ribozymes, e.g. as described in Atkins et al. 1994 (WO9400012), Lenee et al. 1995 (WO9503404), Lutziger et al. 2000 (WO0000619), Prinsen et al. 1997 (WO9713865) and Scott et al. 1997 (WO9738116).

Modulating, including lowering, the level of active gene products or of gene product activity can be achieved by administering or exposing cells, tissues, organs or organisms to said gene product, a homolog, analogue, derivative and/or immunologically active fragment thereof. Immunomodulation is another example of a technique capable of downregulation levels of active gene product and/or of gene product activity and comprises administration of or exposing to or expressing antibodies to said gene product to or in cells, tissues, organs or organisms wherein levels of said gene product and/or gene product activity are to be modulated. Such antibodies comprise "plantibodies", single chain antibodies, IgG antibodies and heavy chain camel antibodies as well as fragments thereof.

Modulating, including lowering, the level of active gene products or of gene product activity can futhermore be achieved by administering or exposing cells, tissues, organs or organisms to an inhibitor or activator of said gene product or the activity thereof. Such inhibitors or activators include proteins (comprising e.g. proteinases and kinases) and chemical compounds identified according to the current invention as described supra.

In the context of the invention the term "agonist" refers to a substance that can be either a protagonist or an antagonist, i.e. can have either positive or negative effects, can be an enhancer or an inhibitor or a modulator as well.

In the context of the current invention is envisaged the downregulation of the expression of a LRR receptor-like kinase gene as defined higher. The invention further comprises downregulation of levels of a LRR receptor-like kinase or of a LRR receptor-like kinase activity whereby said LRR receptor-like kinase has been defined supra.

By "cell fate and/or plant development and/or plant morphology and/or biochemistry and/or physiology" is meant that one or more developmental and/or morphological and/or biochemical and/or physiological characteristics of a plant is altered by the performance of one or more steps pertaining to the invention described herein.

"Cell fate" refers to the cell-type or cellular characteristics of a particular cell that are produced during plant development or a cellular process therefor, in particular during the cell cycle or as a consequence of a cell cycle process.

"Plant development" or the term "plant developmental characteristic" or similar term shall, when used herein, be taken to mean any cellular process of a plant that is involved in determining the developmental fate of a plant cell, in particular the specific tissue or organ type into which a progenitor cell will develop. Cellular processes relevant to plant development will be known to those skilled in the art. Such processes include, for example, morphogenesis, photomorphogenesis, shoot development, root development, vegetative development, reproductive development, stem elongation, flowering, and regulatory mechanisms involved in determining cell fate, in particular a process or regulatory process involving the cell cycle.

"Plant morphology" or the term "plant morphological characteristic" or similar term will, when used herein, be understood by those skilled in the art to refer to the external appearance of a plant, including any one or more structural features or combination of structural features thereof. Such structural features include the shape, size, number, position, colour, texture, arrangement, and patternation of any cell, tissue or organ or groups of cells, tissues or organs of a plant, including the root, stem, leaf, shoot, petiole, trichome, flower, petal, stigma, style, stamen, pollen, ovule, seed, embryo, endosperm, seed coat, aleurone, fibre, fruit, cambium, wood, heartwood, parenchyma, aerenchyma, sieve element, phloem or vascular tissue, amongst others.

"Plant biochemistry" or the term "plant biochemical characteristic" or similar term will, when used herein, be understood by those skilled in the art to refer to the metabolic and catalytic processes of a plant, including primary and secondary metabolism and the products thereof, including any small molecules, macromolecules or chemical compounds, such as but not limited to starches, sugars, proteins, peptides, enzymes, hormones, growth factors, nucleic acid molecules, celluloses, hemicelluloses, calloses, lectins, fibres, pigments such as anthocyanins, vitamins, minerals, micronutrients, or macronutrients, that are produced by plants.

"Plant physiology" or the term "plant physiological characteristic" or similar term will, when used herein, be understood to refer to the functional processes of a plant, including developmental processes such as growth, expansion and differentiation, sexual development, sexual reproduction, seed set, seed development, grain filling, asexual reproduction, cell division, dormancy, germination, light adaptation, photosynthesis, leaf expansion, fiber production, secondary growth or wood production, amongst others; responses of a plant to externally-applied factors such as metals, chemicals, hormones, growth factors, environment and environmental stress factors (eg. anoxia, hypoxia, high temperature, low temperature, dehydration, light, daylength, flooding, salt, heavy metals, amongst others), including adaptive responses of plants to said externally-applied factors.

Means for introducing recombinant DNA into plant tissue or cells include, but are not limited to, transformation using $CaCl_2$ and variations thereof, in particular the method described by Hanahan (1983), direct DNA uptake into protoplasts (Krens *et al*, 1982; Paszkowski *et al*, 1984), PEG-mediated uptake to protoplasts (Armstrong *et al*, 1990) microparticle bombardment, electroporation (Fromm et al., 1985), microinjection of DNA (Crossway et al., 1986), microparticle bombardment of tissue explants or cells (Christou *et al*. 1988), vacuum-infiltration of tissue with nucleic acid, or in the case of plants, T-DNA-mediated transfer from *Agrobacterium* to the plant tissue as described essentially by An et al.(1985), Dodds et al. (1985), Herrera-Estrella et al. (1983a, 1983b). Methods for transformation of monocotyledonous plants are well known in the art and include *Agrobacterium*-mediated transformation (Cheng et al. 1997 - WO9748814; Hansen 1998 - WO9854961, Hiei et al. 1994 - WO9400977; Hiei et al. 1998 - WO9817813; Rikiishi et al 1999 - WO9904618; Saito et al. 1995 - WO9506722), microprojectile bombardment (Adams et al. 1999 - US5969213; Bowen et al. 1998 - US5736369; Chang et al. 1994 - WO9413822; Lundquist et al. 1999 - US5874265/US5990390; Vasil and Vasil 1995 - US5405765; Walker et al. 1999 - US5955362), DNA uptake (Eyal et al. 1993 - WO9318168), microinjection of *Agrobacterium* cells (von Holt 1994 - DE4309203) and sonication (Finer et al. 1997 - US5693512).

For microparticle bombardment of cells, a microparticle is propelled into a cell to produce a transformed cell. Any suitable ballistic cell transformation methodology and apparatus can be used in performing the present invention. Exemplary apparatus and procedures are disclosed by Stomp et al. (U.S. Patent No. 5122466) and Sanford and Wolf (U.S.

Patent No. 4945050). When using ballistic transformation procedures, the gene construct may incorporate a plasmid capable of replicating in the cell to be transformed. Examples of microparticles suitable for use in such systems include 1 to 5 µm gold spheres. The DNA construct may be deposited on the microparticle by any suitable technique, such as by precipitation.

A whole plant may be regenerated from the transformed or transfected cell, in accordance with procedures well known in the art. Plant tissue capable of subsequent clonal propagation, whether by organogenesis or embryogenesis, may be transformed with a gene construct of the present invention and a whole plant regenerated therefrom. The particular tissue chosen will vary depending on the clonal propagation systems available for, and best suited to, the particular species being transformed. Exemplary tissue targets include leaf disks, pollen, embryos, cotyledons, hypocotyls, megagametophytes, callus tissue, existing meristematic tissue (e.g., apical meristem, axillary buds, and root meristems), and induced meristem tissue (e.g., cotyledon meristem and hypocotyl meristem).

The term "organogenesis", as used herein, means a process by which shoots and roots are developed sequentially from meristematic centers.

The term "embryogenesis", as used herein, means a process by which shoots and roots develop together in a concerted fashion (not sequentially), whether from somatic cells or gametes.

Preferably, the plant is produced according to the inventive method is transfected or transformed with a genetic sequence, or amenable to the introduction of a protein, by any art-recognized means, such as microprojectile bombardment, microinjection, *Agrobacterium*-mediated transformation (including the 'flower dip' transformation method; Bechtold and Pelletier 1998, Trieu et al. 2000), protoplast fusion, or electroporation, amongst others. Most preferably said plant is produced by *Agrobacterium*-mediated transformation.

The "seedling" is the juvenile plant that arises from the mature embryo after seed germination.

With "differentiation of a cell" it is understood that the cell develops unique features to be engaged for a specific function. Mostly differentiation is irreversible.

Agrobacterium-mediated transformation or agrolistic transformation of plants, yeast, moulds or filamentous fungi is based on the transfer of part of the transformation vector

sequences, called the T-DNA, to the nucleus and on integration of said T-DNA in the genome of said eukaryote.

With "*Agrobacterium*" is meant a member of the *Agrobacteriaceae*, more preferably *Agrobacterium* or *Rhizobacterium* and most preferably *Agrobacterium tumefaciens*.

With "T-DNA", or transferred DNA, is meant that part of the transformation vector flanked by T-DNA borders which is, after activation of the *Agrobacterium vir* genes, nicked at the T-DNA borders and is transferred as a single stranded DNA to the nucleus of an eukaryotic cell.

When used herein, with "T-DNA borders", "T-DNA border region", or "border region" are meant either right T-DNA border (RB) or left T-DNA border (LB). Such a border comprises a core sequence flanked by a border inner region as part of the T-DNA flanking the border and/or a border outer region as part of the vector backbone flanking the border. The core sequences comprise 22 bp in case of octopine-type vectors and 25 bp in case of nopaline-type vectors. The core sequences in the right border region and left border region form imperfect repeats. Border core sequences are indispensable for recognition and processing by the *Agrobacterium* nicking complex consisting of at least VirD1 and VirD2. Core sequences flanking a T-DNA are sufficient to promote transfer of said T-DNA. However, efficiency of transformation using transformation vectors carrying said T-DNA solely flanked by said core sequences is low. Border inner and outer regions are known to modulate efficiency of T-DNA transfer (Wang et al. 1987). One element enhancing T-DNA transfer has been characterized and resides in the right border outer region and is called *overdrive* (Peralta et al. 1986, van Haaren et al. 1987).

With "T-DNA transformation vector" or "T-DNA vector" is meant any vector encompassing a T-DNA sequence flanked by a right and left T-DNA border consisting of at least the right and left border core sequences, respectively, and used for transformation of any eukaryotic cell.

With "T-DNA vector backbone sequence" or "T-DNA vector backbone sequences" is meant all DNA of a T-DNA containing vector that lies outside of the T-DNA borders and, more specifically, outside the nicking sites of the border core imperfect repeats.

The current invention includes optimized T-DNA vectors such that vector backbone integration in the genome of a eukaryotic cell is minimized or absent. With "optimized T-DNA vector" is meant a T-DNA vector designed either to decrease or abolish transfer of vector backbone sequences to the genome of a eukaryotic cell. Such T-DNA vectors are

known to the one familiar with the art and include those described by Hanson et al. (1999) and by Stuiver et al. (1999 - WO9901563).

The current invention clearly considers the inclusion of a DNA sequence comprising the promoter sequence of the present invention encoding a LRR receptor-like kinase, homolog, derivative or immunologically active fragment thereof as defined supra, in any T-DNA vector comprising binary transformation vectors, super-binary transformation vectors, co-integrate transformation vectors, Ri-derived transformation vectors as well as in T-DNA carrying vectors used in agrolistic transformation.

With "binary transformation vector" is meant a T-DNA transformation vector comprising: a T-DNA region comprising at least one gene of interest and/or at least one selectable marker active in the eukaryotic cell to be transformed; and

a vector backbone region comprising at least origins of replication active in *E. coli* and *Agrobacterium* and markers for selection in *E. coli* and *Agrobacterium*.

The T-DNA borders of a binary transformation vector can be derived from octopine-type or nopaline-type Ti plasmids or from both. The T-DNA of a binary vector is only transferred to a eukaryotic cell in conjunction with a helper plasmid. Also known in the art are multiple binary vector *Agrobacterium* strains for efficient co-transformation of plants (Bidney and Scelonge 2000 - WO0018939).

With "helper plasmid" is meant a plasmid that is stably maintained in *Agrobacterium* and is at least carrying the set of *vir* genes necessary for enabling transfer of the T-DNA. Said set of *vir* genes can be derived from either octopine-type or nopaline-type Ti plasmids or from both.

With "super-binary transformation vector" is meant a binary transformation vector additionally carrying in the vector backbone region a *vir* region of the Ti plasmid pTiBo542 of the super-virulent *A. tumefaciens* strain A281 (Hiei et al. 1994 - EP0604662, Hiei et al. 1995 - EP0687730). Super-binary transformation vectors are used in conjunction with a helper plasmid.

With "co-integrate transformation vector" is meant a T-DNA vector at least comprising: a T-DNA region comprising at least one gene of interest and/or at least one selectable marker active in plants; and

a vector backbone region comprising at least origins of replication active in *Escherichia coli* and *Agrobacterium*, and markers for selection in *E. coli* and *Agrobacterium*, and a set of *vir* genes necessary for enabling transfer of the T-DNA.

The T-DNA borders and said set of *vir* genes of a said T-DNA vector can be derived from either octopine-type or nopaline-type Ti plasmids or from both.

With "Ri-derived plant transformation vector" is meant a binary transformation vector in which the T-DNA borders are derived from a Ti plasmid and said binary transformation vector being used in conjunction with a 'helper' Ri-plasmid carrying the necessary set of *vir* genes.

As used herein, the term "selectable marker gene" or "selectable marker" or "marker for selection" includes any gene which confers a phenotype on a cell in which it is expressed to facilitate the identification and/or selection of cells which are transfected or transformed with a gene construct of the invention or a derivative thereof. Suitable selectable marker genes contemplated herein include the ampicillin resistance (Amp^r), tetracycline resistance gene (Tc^r), bacterial kanamycin resistance gene (Kan^r), phosphinothricin resistance gene, neomycin phosphotransferase gene (*npt*II), hygromycin resistance gene, β-glucuronidase (GUS) gene, chloramphenicol acetyltransferase (CAT) gene, green fluorescent protein (*gfp*) gene (Haseloff *et al*, 1997), and luciferase gene, amongst others.

With "agrolistics", "agrolistic transformation" or "agrolistic transfer" is meant here a transformation method combining features of *Agrobacterium*-mediated transformation and of biolistic DNA delivery. As such, a T-DNA containing target plasmid is co-delivered with DNA/RNA enabling in planta production of VirD1 and VirD2 with or without VirE2 (Hansen and Chilton 1996; Hansen et al. 1997; Hansen and Chilton 1997 - WO9712046). The present invention further describes an approach to remove from transformed cells a stably integrated foreign DNA sequence by recombination involving a recombinase and recombination sites.

With "foreign DNA" is meant any DNA sequence that is introduced in the host's genome by recombinant techniques. Said foreign DNA includes e.g. a T-DNA sequence or a part thereof such as the T-DNA sequence comprising the selectable marker in an expressible format. Foreign DNA furthermore include intervening DNA sequences as defined supra.

With "recombination event" is meant either a site-specific recombination event or a recombination event effected by transposon 'jumping'.

With "recombinase" is meant either a site-specific recombinase or a transposase.

With "recombination site" is meant either site-specific recombination sites or transposon border sequences.

With "site specific recombination event" is meant an event catalyzed by a system generally consisting of three elements: a pair of DNA sequences (the site-specific recombination sequences or sites) and a specific enzyme (the site-specific recombinase). The site-specific recombinase catalyzes a recombination reaction only between two site-specific recombination sequences depending on the orientation of the site-specific recombination sequences. Sequences intervening between two site-specific recombination sites will be inverted in the presence of the site-specific recombinase when the site-specific recombination sequences are oriented in opposite directions relative to one another (i.e. inverted repeats). If the site-specific recombination sequences are oriented in the same direction relative to one another (i.e. direct repeats), then any intervening sequences will be deleted upon interaction with the site-specific recombinase. Thus, if the site-specific recombination sequences are present as direct repeats at both ends of a foreign DNA sequence integrated into a eukaryotic genome, such integration of said sequences can subsequently be reversed by interaction of the site-specific recombination sequences with the corresponding site specific recombinase. A number of different site specific recombinase systems can be used including but not limited to the Cre/lox system of bacteriophage P1, the FLP/FRT system of yeast, the Gin recombinase of phage Mu, the Pin recombinase of *E. coli*, the PinB, PinD and PinF from *Shigella*, and the R/RS system of *Zygosaccharomyces rouxii*. Recombinases generally are integrases, resolvases or flippases. Also dual-specific recombinases can be used in conjunction with direct or indirect repeats of two different site-specific recombination sites corresponding to the dual-specific recombinase (Baszczynski et al. 1999 - WO9925840). The preferred site-specific recombinase systems are the bacteriophage P1 Cre/lox, the yeast FLP/FRT and the *Z. rouxii* R/RS systems. In these systems a recombinase (Cre, FLP or R) interact specifically with its respective site-specific recombination sequence (lox, FRT, or RS respectively) to invert or excise the intervening sequences. The site-specific recombination sequences for each of these two systems are relatively short (34 bp for lox and 47 bp for FRT). Some of these systems have already been used with high efficiency in plants such as tobacco (Dale et al. 1990, Onouchi et al. 1991, Sugita et al. 2000) and *Arabidopsis* (Osborne et al. 1995, Onouchi et al. 1995). Site-specific recombination systems have many applications in plant molecular biology including methods for control of homologous recombination (e.g. Hodges et al. 1996 - US5527695), for targeted insertion, gene stacking, etc.

(Baszczynski et al. 1999 - WO9925821) and for resolution of complex T-DNA integration patterns or for excision of a selectable marker (Ow et al. 1999 - WO9923202).

Although the site-specific recombination sequences must be linked to the ends of the DNA to be excised or to be inverted, the gene encoding the site specific recombinase may be located elsewhere. For example, the recombinase gene could already be present in the eukaryote's DNA or could be supplied by a later introduced DNA fragment either introduced directly into cells, through crossing or through cross-pollination. Alternatively, a substantially purified recombinase protein could be introduced directly into the eukaryotic cell, e.g. by micro-injection or particle bombardment. Typically, the site-specific recombinase coding region will be operably linked to regulatory sequences enabling expression of the site-specific recombinase in the eukaryotic cell.

With "recombination event effected by transposon jumping" or "transposase-mediated recombination" is meant a recombination event catalyzed by a system consisting of three elements: a pair of DNA sequences (the transposon border sequences) and a specific enzyme (the transposase). The transposase catalyzes a recombination reaction only between two transposon border sequences which are arranged as inverted repeats. A number of different transposon/transposase systems can be used including but not limited to the Ds/Ac system, the Spm system and the Mu system. These systems originate from corn but it has been shown that at least the Ds/Ac and the Spm system also function in other plants (Fedoroff et al. 1993, Schlappi et al. 1993, Van Sluys et al. 1987). Preferred are the Ds- and the Spm-type transposons which are delineated by 11 bp- and 13 bp- border sequences, respectively.

Although the transposon border sequences must be linked to the ends of the DNA to be excised, the gene encoding the transposase may be located elsewhere. For example, the recombinase gene could already be present in the eukaryote's DNA or could be supplied by a later introduced DNA fragment either introduced directly into cells, through crossing or through cross-pollination. Alternatively, a substantially purified transposase protein could be introduced directly into cells, e.g. by microinjection or by particle bombardment.

As part of the current invention, transposon border sequences are included in a foreign DNA sequence such that they lie outside said DNA sequence and transform said DNA into a transposon-like entity that can move by the action of a transposase.

As transposons often reintegrate at another locus of the host's genome, segregation of the progeny of the hosts in which the transposase was allowed to act might be necessary to separate transformed hosts containing e.g. only the transposon footprint and transformed hosts still containing the foreign DNA.

In performing the present invention, the genetic element is preferably induced to mobilize, such as, for example, by the expression of a recombinase protein in the cell which contacts the integration site of the genetic element and facilitates a recombination event therein, excising the genetic element completely, or alternatively, leaving a "footprint", generally of about 20 nucleotides in length or greater, at the original integration site. Those hosts and host parts that have been produced according to the inventive method can be identified by standard nucleic acid hybridization and/or amplification techniques to detect the presence of the mobilizable genetic element or a gene construct comprising the same. Alternatively, in the case of transformed host cells, tissues, and hosts wherein the mobilizable genetic element has bene excised, it is possible to detect a footprint in the genome of the host which has been left following the excision event, using such techniques. As used herein, the term "footprint" shall be taken to refer to any derivative of a mobilizable genetic element or gene construct comprising the same as described herein which is produced by excision, deletion or other removal of the mobilizable genetic element from the genome of a cell transformed previously with said gene construct. A footprint generally comprises at least a single copy of the recombination loci or transposon used to promote excision. However, a footprint may comprise additional sequences derived from the gene construct, for example nucleotide sequences derived from the left border sequence, right border sequence, origin of replication, recombinase-encoding or transposase-encoding sequence if used, or other vector-derived nucleotide sequences. Accordingly, a footprint is identifiable according to the nucleotide sequence of the recombination locus or transposon of the gene construct used, such as, for example, a sequence of nucleotides corresponding or complementary to a *lox* site, *frt* site or RS site.

The term "cell cycle" means the cyclic biochemical and structural events associated with growth and with division of cells, and in particular with the regulation of the replication of DNA and mitosis. Cell cycle includes phases called: G0, Gap1 (G1), DNA synthesis (S), Gap2 (G2), and mitosis (M). Normally these four phases occur sequentially, however, the cell cycle also includes modified cycles wherein one or more phases are absent

resulting in modified cell cycle such as endomitosis, acytokinesis, polyploidy, polyteny, and endoreduplication.

The term "cell cycle interacting protein", "cell cycle protein" or "cell cycle control protein" as denoted herein means a protein which exerts control on or regulates or is required for the cell cycle or part thereof of a cell, tissue, organ or whole organism and/or DNA replication. It may also be capable of binding to, regulating or being regulated by cyclin dependent kinases or their subunits. The term also includes peptides, polypeptides, fragments, variant, homologs, alleles or precursors (eg preproproteins or preproteins) thereof.

Cell cycle control proteins and their role in regulating the cell cycle of eukaryotic organisms are reviewed in detail by John (1981) and the contributing papers therein (Norbury & Nurse 1992;Nurse 1990;Ormrod & Francis 1993) and the contributing papers therein (Doerner et al. 1996) which are incorporated by reference.(Elledge 1996;Francis, Francis & Halford 1995;Hirt et al. 1991;Mironov et al. 1999)

The term "cell cycle control genes" refers to any gene or mutant thereof which exerts control on or are required for: chromosomal DNA synthesis and for mitosis (preprophase band, nuclear envelope, spindle formation, chromosome condensation, chromosome segregation, formation of new nuclei, formation of phragmoplast, duplication of microtubule-organizing center, etc) meiosis, cytokinesis, cell growth, endoreduplication, cell cycle control genes are also all genes exerting control on the above: homologs of CDKs, cyclins, E2Fs, Rb, CKI, Cks, and also any genes which interfere with the above, cyclin D, cdc25, Wee1, Nim1, MAP kinases, etc.

More specifically, cell cycle control genes are all genes involved in the control of entry and progression through S phase. They include, not exclusively, genes expressing "cell cycle control proteins" such as cyclin dependent kinases (CDK), cyclin dependent kinase inhibitors (CKI), D, E and A cyclins, E2F and DP transcription factors, pocket proteins, CDC7/DBF4 kinase, CDC6, MCM2-7, Orc proteins, cdc45, components of SCF ubiquitin ligase, PCNA, DNA-polymerase.

The term "cell cycle control protein" include cyclins A, B, C, D and E including CYCA1;1, CYCA2;1, CYCA3;1, CYCB1;1, CYCB1;2, CYC B2;2, CYCD1;1, CYCD2;1, CYCD3;1, and CYCD4;1 (Evans et al. 1983;Francis et al. 1998;Labbe et al. 1989;Murray & Kirschner 1989;Renaudin et al. 1996;Soni et al. 1995;Sorrell et al. 1999;Swenson, Farrell, & Ruderman 1986) cyclin dependent kinase inhibitor (CKI) proteins such as ICK1

(Wang, Fowke, & Crosby 1997), FL39, FL66, FL67 (PCT/EP98/05895), Sic1, Far1, Rum1, p21, p27, p57, p16, p15, p18, p19 (Elledge 1996;Pines 1995), p14 and p14ARF; p13suc1 or CKS1At (De Veylder et al. 1997;Hayles & Nurse 1986) and nim-1 (Fantes 1989;Russell & Nurse 1986;Russell & Nurse 1987b;Russell & Nurse 1987a;Russell & Nurse 1986) homologues of Cdc2 such as Cdc2MsB (Hirt et al. 1993) CdcMs kinase (Bogre et al. 1997) cdc2 T14Y15 phosphatases such as Cdc25 protein phosphatase or p80cdc25 (Bell et al. 1993;Elledge 1996;Kumagai & Dunphy 1991;Russell et al. 1986) and Pyp3 (Elledge 1996) cdc2 protein kinase or p34cdc2 (Colasanti, Tyers, & Sundaresan 1991;Feiler & Jacobs 1990;Hirt et al. 1991;John, Sek, & Lee 1989;Lee & Nurse 1987;Nurse & Bissett 1981;Ormrod et al. 1993) cdc2a protein kinase (Hemerly et al. 1993) cdc2 T14Y15 kinases such as wee1 or p107wee1 (Elledge 1996;Russell et al. 1986;Russell et al. 1987b;Russell et al. 1987a;Sun et al. 1999) mik1 (Lundgren et al. 1991) and myt1 (Elledge 1996); cdc2 T161 kinases such as Cak and Civ (Elledge 1996); cdc2 T161 phosphatases such as Kap1 (Elledge 1996); cdc28 protein kinase or p34cdc28 (Nasmyth 1993;Reed, Hadwiger, & Lorincz 1985) p40MO15 (Fesquet et al. 1993;Poon et al. 1993) chk1 kinase (Zeng et al. 1998) cds1 kinase (Zeng et al. 1998) growth-associated H1 kinase (Labbe et al. 1989;Lake & Salzman 1972;Langan 1978;Zeng et al. 1998) MAP kinases described by (Binarova et al. 1998;Bögre et al. 1999;Calderini et al. 1998;Wilson et al. 1999).

Other cell cycle control proteins that are involved in cyclin D-mediated entry of cells into G1 from G0 include pRb (Huntley et al. 1998;Xie et al. 1996) E2F, RIP, MCM7 and potentially the pRb-like proteins p107 and p130.

Other cell cycle control proteins that are involved in the formation of a pre-replicative complex at one or more origins of replication, such as, but not limited to, ORC, CDC6, CDC14, RPA and MCM proteins or in the regulation of formation of this pre-replicative complex, such as, but not limited to, the CDC7, DBF4 and MBF proteins.

For the present purpose, the term "cell cycle control protein" shall further be taken to include any one or more of those proteins that are involved in the turnover of any other cell cycle control protein, or in regulating the half-life of said other cell cycle control protein. The term "protein turnover" is to include all biochemical modifications of a protein leading to the physical or functional removal of said protein. Although not limited to these, examples of such modifications are phosphorylation, ubiquitination and proteolysis. Particularly preferred proteins which are involved in the proteolysis of one or

more of any other of the above-mentioned cell cycle control proteins include the yeast-derived and animal-derived proteins, Skp1, Skp2, Rub1, Cdc20, cullins, CDC23, CDC27, CDC16, and plant-derived homologues thereof (Cohen-Fix & Koshland 1997;Hochstrasser 1998;Krek 1998;Lisztwan et al. 1998) and Plesse et al in (Francis et al. 1998)).

For the present purpose, the term "cell cycle control genes" shall further be taken to include any one or more of those gene that are involved in the transcriptional regulation of cell cycle control gene expression such as transcription factors and upstream signal proteins. Additional cell cycle control genes are not excluded.

For the present purpose, the term "cell cycle control genes" shall further be taken to include any cell cycle control gene or mutant thereof, which is affected by environmental signals such as for instance stress, nutrients, pathogens, or by intrinsic signals such as the animal mitogens or the plant hormones (auxins, cytokinins, ethylene, gibberellic acid, abscisic acid and brassinosteroids).

The term "cell cycle progression" refers to the process of passing through the different cell cycle phases. The term "cell cycle progression rate" accordingly refers to the speed at which said cell cycle phases are run through or the time spans required to complete said cell cycle phases.

With "pathogen" is meant those organisms that have a negative effect on the physiological state of the plant or a part thereof. Some pathogens are for instance viruses, bacteria, fungi, and parasitic plants. with plant "pests" is meant the group of nematodes as well as insects, which are able to attack and to damage the plant.

"Plant cell" comprises any cell derived from any plant and existing in culture as a single cell, a group of cells or a callus. A plant cell may also be any cell in a developing or mature plant in culture or growing in nature.

"Plants" comprises all plants, including monocotyledonous and dicotyledonous plants.

With "yeast two-hybrid assay" is meant an assay that is based on the observation that many eukaryotic transcription factors comprise two domains, a DNA-binding domain (DB) and an activation domain (AD) which, when physically separated (i.e. disruption of the covalent linkage) do not effectuate target gene expression. Two proteins able to interact physically with one of said proteins fused to DB and the other of said proteins fused to AD will re-unite the DB and AD domains of the transcription factor resulting in target gene expression. The target gene in the yeast two-hybrid assay is usually a

reporter gene such as the β-galactosidase gene. Interaction between protein partners in the yeast two-hybrid assay can thus be quantified by measuring the activity of the reporter gene product (Bartel and Fields 1997). Alternatively, a mammalian two-hybrid system can be used which includes e.g. a chimeric green fluorescent protein encoding reporter gene (Shioda et al. 2000). Yet another alternative consists of a bacterial two-hybrid system using e.g. *HIS* as reporter gene (Joung et al. 2000).

The term "fragment of a sequence" or "part of a sequence" means a truncated sequence of the original sequence referred to. The truncated sequence (nucleic acid or protein sequence) can vary widely in length; the minimum size being a sequence of sufficient size to provide a sequence with at least a comparable function and/or activity or the original sequence referred to, while the maximum size is not critical. In some applications, the maximum size usually is not substantially greater than that required to provide the desired activity and/or function(s) of the original sequence. Typically, the truncated amino acid or nucleotide sequence will range from about 5 to about 60 amino acids in length. More typically, however, the sequence will be a maximum of about 50 amino acids in lenght, preferably a maximum of about 60 amino acids. It is usually desirable to select sequences of at least about 10, 12 or 15 amino acids or nucleotides, up to a maximum of about 20 or 25 amino acids or nucleotides.

Furthermore, folding simulations and computer redesign of structural motifs of the protein of the invention can be performed using appropriate computer programs (Olszewski et al.1996, Hoffman et al.1995). Computer modeling of protein folding can be used for the conformational and energetic analysis of detailed peptide and protein models (Monge et al. 1995, Renouf et al. 1995). In particular, the appropriate programs can be used for the identification of interactive sites of the LRR receptor-like kinase of the present invention by computer assistant searches for complementary peptide sequences (Fassina and Melli 1994). Further appropriate computer systems for the design of protein and peptides are described in the prior art, for example in Berry and Brenner (1994), Wodak (1987), Pabo and Suchanek (1986). The results obtained form the above-described computer analysis can be used for, e.g. the preparation of peptidomimetics of the protein of the invention or fragments thereof. Such pseudopeptide analogues of the natural amino acid sequence of the protein may very efficiently mimic the parent protein (Benkirane et al. 1996). For example, incorporation of easily available achiral ω-amino acid residues into a protein of the invention or a

fragment thereof results in the substitution of amino bonds by polymethylene units of an aliphatic chain, thereby providing a convenient strategy for constructing a peptidomimetic (Banerjee et al. 1996). Superactive peptidomimetic analogues of small peptide hormones in other systems are described in the prior art (Zhang et al. 1996). Appropriate peptidomimetics of the protein of the present invention can also be identified by the synthesis of peptidomimetic combinatorial libraries through successive amine alkylation and testing the resulting compounds, e.g., for their binding, kinase inhibitory and/or immunlogical properties. Methods for the generation and use of peptidomimetic combinatorial libraries are described in the prior art, for example in Ostresh et al. (1996) and Dorner et al. (1996).

Furthermore, a three-dimensional and/or crystallographic structure of the protein of the invention can be used for the design of peptidomimetic inhibitors of the biological activity of the protein of the invention (Rose et al. 1996, Ruterber et al. 1996).

To align sequences we used the Clustal V method described in Higgins and Sharp (Higgins DG & Sharp PM, *Comput Appl Biosci.* 1989 Apr;5(2)151-3). The Clustal Method groups sequences into clusters by examining sequence distances between all pairs. Clusters are aligned as pairs then collectively as sequence groups to produce the overall alignment. After the multiple alignment is completed, a neighbor-joining method is employed to reconstruct phylogeny for the putative alignment.

1. Residue weight table

The Clustal method uses weight tables to construct multiple alignments. Residue weight tables are used in scoring protein and nucleotide alignments so that slightly mismatched residues, such as Ile vs. Val amino acids or G vs. R nucleotides, are not scored the same as total mismatches.

If your sequences are protein, or a mixture of protein and DNA, choose between PAM250, PAM100, Structural and Identity.

1. iPAM250 is recommended for the Clustal method and is ideal for longer sequences or highly diverged sequences.
2. PAM100 is also used with the Clustal method and is most accurate for shorter sequences or sequences having little divergence.
3. iStructural is recommended for the Jotun Hein method and gives higher scores to residues that are chemically and spatially similar.

4. Identity may be used with Clustal or Jotun Hein methods, but is recommended only for the Clustal method. The table scores non-identities as zero, and identities as a default value of 10.

2. Method Parameters

A. Pairwise alignment parameters may include K-tuple, Gap Penalty, Window and Diagonals Saved:

1. Specify the K-tuple, expressed as the unit's length in nucleotides. Any stretch of K or more adjacent nucleotides that form an exact match in both sequences is a K-tuple. A lower K-tuple will find smaller matching regions; a higher value will find fewer, but better, matching regions of similarity.
2. Type in the Gap Penalty, the amount deducted from the score for each gap in the alignment. Gaps of different sizes carry the same penalty.
3. Specify the Window, the range the program searches for all possible or alternate K-tuples. An extremely high value may slow an alignment by considering subalignments far from the main diagonal.
4. Type in the number of Diagonals Saved. In the pairwise calculation, the best scoring diagonals for each pair in the matrix are saved and used in the clustering and scoring of ancestors.

B. Multiple alignment parameters include Gap Penalty and Gap Length Penalty.

1. Specify the Gap Penalty, the amount deducted from the score for each gap in the alignment. Gaps of different sizes carry the same penalty.
2. Specify the Gap Length Penalty, the value which will be deducted from the alignment score after first multiplying it by the number of gaps. Longer gaps have a greater penalty than shorter gaps.

Whenever in the current description reference is made to a "phylogenetic tree", this phylogenetic tree is used to view phylogenetic relationships compatible with multiple sequence alignments. The length of each pair of branches represents the distance between sequence pairs, while the units at the bottom of the tree indicate the number of substitution events.

The phylogenetic tree gives a first estimate of the relationships between homologous sequences. It is highly recommended that your data be evaluated using additional methods before final conclusions are made about evolutionary relationships.

You can switch the view between Balanced Branches (phenogram) or Unbalanced Branches (cladogram). A phenogram averages the distances between ancestors in the tree. Dotted lines indicate a negative branch length, which may be a byproduct of averaging. In the cladogram, branch distances correspond to sequence divergence.

The compounds to be obtained or identified in the methods of the invention can be compounds that are able to bind to any of the nucleic acids, peptides or proteins of the invention. Other interesting compounds to be identified are compounds that modulate the expression of the genes or the proteins of the invention in such a way that either the expression of said gene or protein is enhanced or decreased by the action of said compound. Alternatively the compound can exert his action by directly or indirectly enhancing or decreasing the activity of any of the proteins of the invention.

Said compound or plurality of compounds may be comprised in, for example, samples, e.g., cell extracts from, e.g., plants, animals or microorganisms. Furthermore, said compound(s) may be known in the art but hitherto not known to be capable of suppressing or activating cell cycle interacting proteins. The reaction mixture may be a cell free extract of may comprise a cell or tissue culture. Suitable set ups for the method of the invention are known to the person skilled in the art and are, for example, generally described in Alberts et al., Molecular Biology of the Cell, third edition (1994), in particular Chapter 17. The plurality of compounds may be, e.g., added to the reaction mixture, culture medium or injected into the cell.

If a sample containing a compound or a plurality of compounds is identified in the method of the invention, then it is either possible to isolate the compound form the original sample identified as containing the compound capable of acting as an agonist, or one can further subdivide the original sample, for example, if it consists of a plurality of different compounds, so as to reduce the number of different substances per sample and repeat the method with the subdivisions of the original sample. Depending on the complexity of the samples, the steps described above can be performed several times, preferably until the sample identified according to the method of the invention only comprises a limited number of or only one substance(s). Preferably said sample comprises substances or similar chemical and/or physical properties, and most

preferably said substances are identical. Preferably, the compound identified according to the above described method or its derivative is further formulated in a form suitable for the application in plant breeding or plant cell and tissue culture.

The invention, now being generally described, will be more readily understood by reference to the following examples, which are included merely for purposes of illustration of certain aspects and embodiments of the present invention and are not intended to limit the invention.

DESCRIPTION OF FIGURES

Figure 1. Models for stem-cell maintenance in the root and shoot meristem.
(a) In the root meristem, the quiescent center (QC) functions as an organizing center. Those cells in immediate contact with the QC are stem cells (dark shading), which are initials for all cell files: endodermis (end), cortex (cort), epidermis (epi), lateral root cap (lrc) and columella (col). (b) In the shoot meristem, WUS is functionally equivalent to the QC in the root meristem in maintaining the presence of stem cells (dark shading). Long-range signaling for stem cell maintenance is also postulated for ZLL, which is expressed in the vascular cells. CLV3 is expressed in the presumptive stem cells that the initials for the L1 (epidermis), L2 (subepidermal tissues, i.e. mesophyll cells) and L3 (subepidermal tissues, i.e. vascular cells) tissue layers (Schnittger, Schellmann, & Hulskamp 1999). CLV1 is found in the L3 tunica layer of the central zone.

Figure 2. List of the degenerative primers used for RT-PCR of root-specific members of the LRR receptor kinase family.

Figure 3. Differential RT-PCR on hypocotyl/cotyledon (C) and roottip (R) RNA.
The PCR fragments obtained upon amplification of R and C cDNAs with degenerate primers are separated on agarose gel. DNA fragments of different size can be detected in the R_{22} compared to C_{22} PCRs.
M represents the 100 bp marker (Boehringer). C and R indicated the origin of the cDNA sample used in the PCR and the numbers below indicate the primer combination used, e.g. C_{23} represents hypocotyl/cotyledon cDNA amplified with primers LRR2 and KIN3.

Figure 4. RT - PCR analysis of *RCH1* expression.
Total RNA was isolated from flower (F), hypocotyl/cotyledon (C) and roottips (R) and *RCH1* mRNA was amplified for 26 cycles with RT-PCR using the RCH1F and RCH1R primers. As a control for cDNA quality and quantity as well as relative expression level, ubiquitin (Ubi) mRNA was amplified for 20 cycles under the same conditions using the UBIF and UBIR primers.

Figure 5. The nucleotide sequence of the RCH1 operon.

This figure shows het nucleotide sequence of SEQ ID NO 1 that contains the genomic RCH1 clone.

Figure 6. The nucleotide sequence of the RCH1 promoter region.

This fragment of the genomic RCH1 clone is about 3498 base pairs and is identified as SEQ ID NO 2. Two putative ATG start codons of the RCH1 protein on position 3479-3481 and on position 3491-3493 could be identified and are represented in bold. The underlined sequence corresponds to SEQ ID NO 18.

It is also known that there is a gene in the reverse orientation in this promoter region. This gene's open reading frame starts at the position 1759-1757 with the ATG (bold), contains two introns, and ends at position 350-348 with a TGA codon (bold). A homology search of the protein sequence of this antisense gene in the RCH1prom against the protein database with the program BLASTP 2.0.8 (Altschul et al. 1997) produced a hypothetical protein (Accession number AC002392, clone T20K24.8) that is a putative anthranilate N-hydroxycinnamoyl/benzoyltransferase from *Arabidopsis thaliana*. Both sequences show 38% identity and 54% of their amino acids are homologous.

Figure 7. Promoter activity of the RCH1 promoter – GUS fusion construct in the *Arabidopsis thaliana* roots.

Seedlings of different age were grown and stained with Glucuronidase (GUS). The expression of the GUS reporter gene can be seen in the root meristem of the main and the lateral roots, as well as in the vascular tissue of the root and the lateral root primordia. (A) 1 day old seeding stained over night, (B) higher magnification of A, (C) 2 days old seedling stained for 5 hours, (D) main root from 1 day old seedling stained for 1 hour, (E) lateral root primordia from 1 week old seedling stained for 5 hours, (F, G, H) lateral roots from an 8 days old seedling stained for 5 hours.

Figure 8. The nucleotide sequence of the RCH1 gene.

This fragment of the genomic RCH1 clone is about 3960 base pairs long and is identified as SEQ ID NO 3. Two putative ATG start codons of the RCH1 protein on position 1-3 and on position 13-15 could be identified and are represented in bold. Also the TAA

stopcodon on position 3958-3960 is represented in bold. There is a putative intron estimated to range from position 2900 (CAG^GTA) to position 3452 (CAG^AAT) with the program http://www.CBS.DTU.DK/services/NetGene2. The putative splice-sites are underlined.

Figure 9. The amino acid sequence of the open reading frame encoding RCH1 protein.

Two putative start methionins of the RCH1 protein on position 1 and on position 4 could be identified. An ORF of 1135 amino acids is identified and represented in SEQ ID NO 4. The ORF of 1131 amino acids and starting at the methionine on position 4 (M) corresponds with SEQ ID NO 19.

The boxed peptide RYKIILGAAQGLAYLHHDCVPPIVH corresponds with SEQ ID NO 5.

Figure 10. Amino acid alignment of the RCH1 protein *the Arabidopsis thaliana* CLV1 receptor kinase like protein.

The amino acid sequences of the RCH1 and of CLV1 (Genbank accession number AAB58929) were aligned using the program GAP as described in example 3. Match display thresholds for the alignment(s) are: | = IDENTITY, : = 2, . = 1.

Figure 11. Phylogenetic tree of CLV1 homologues.

The novel RCH1 sequence was aligned with the CLV1 (Genbank accession number U96879.1), the known CLV1 homologues BRI1 (Genbank accession number AF017056), HAE (GenPept accession number S27756) and ER (Genbank accession number D83257) as well as with 4 other related proteins that were retrieved from the public database searched with the RCH1 sequence using the BlastP program. The multiple alignment and phylogenetic tree of said sequences were calculated with the Clustal V method as described in example 4. Based on the sequence, RCH1 is more homologue to CLV1 than other known CLV-homologues.

Figure 12. Phylogenetic tree of LRR receptor like kinases

The novel RCH1 sequence was aligned with CLV1, its homologues BRI, HAE, ER and other LRR receptor like kinases retrieved from the public database. Also plant disease resistance genes, which comprise a leucine-rich repeat and which are receptor-like were

included in the alignment (RPS5, RPS2, RPP8,RPM1, RPS4, N, RPP5, Meyers *et al.*, 1999, *Plant J;*, Nov20(3): 317-32). The multiple alignment and phylogenetic tree of said proteins were calculated with the Clustal V method as described in example 4. Because of the close proximity of RCH1 to CLV1 in the phylogenetic tree, there is more evolutionary relationship between RCH1 and clavata, than with other kinds of LRR receptor like proteins such as plant disease resistance genes.

Figure 13. The nucleotide sequence of a minimal promoter sequence of RCH1

This figure shows the nucleotide sequence of SEQ ID NO. 18 that contains a minimal promoter sequence of the RCH1 clone. This fragment of the genomic RCH1 clone is about 842 base pairs. This sequence is deducted from the large promoter region identified as SEQ ID NO 2 (see example 7). The minimal promoter region correlates with the base pairs ranging from –1 of the first ATG defined in SEQ ID NO. 2 to –842 of the first ATG defined in SEQ ID NO. 2.

Figure 14. Outprint of GenBank accession number AB017061 on December 4, 2000.

Figure 15. RCH1prom::GUS expression in embryo's

Arabidopsis embryo transformed with RCH1prom::GUS. Embryo's were stained with beta glucuronidase. The RCH1 promoter is active very early in plant development and is only active in the root in the embryonic phase (A). On a stronger magnification (B) it is visible that the root specific expression is limited to the meristematic zone of the embryonic root.

Figure 16. RCH1::GFP expression in roots

RCH1 promoter is strongly active in endodermis, cortex, epidermis, lateral root cap, in short: in the "division zone". The RCH1 promoter is also active but in a low manner, in the quiescent center and in the vascular tissue.

Figure 17. Amino acid sequence of RCH1 homologues of Table 4

EXAMPLES

Unless stated otherwise in the Examples, all recombinant DNA techniques are performed according to protocols as described in Sambrook et al. (1989), Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press, NY or in Volumes 1 and 2 of Ausubel et al., (1994), Current Protocols in Molecular Biology, Current Protocols. Standard materials and methods for plant molecular work are described in Plant Molecular Biology Labfase (1993) by R.D.D. Croy, jointly published by BIOS Scientific Publications Ltd. (UK) and Blackwell Scientific Publications (UK).

Example 1: Isolation of the RCH1 operon from root tissue of *Arabidopsis thaliana*.

Plant organ development depends on the presence of distally positioned groups of continuously dividing cells, the shoot and root apical meristems. Laser ablation studies suggest that a balance between signals for proper differentiation and short range signals from the mitotically inactive quiescent centre (QC), required to keep cells less differentiated, is important to maintain the root meristem (van den Berg et al. 1995;van den Berg et al. 1997). In the shoot meristem a balance between the *WUSCHEL* and *CLAVATA* (*CLV*) genes has been implicated to play an essential role in regulation cell division and differentiation (Clark et al. 1993;Fletcher et al. 1999;Schoof et al. 2000). The *CLV1* gene encodes a leucine rich repeat (LRR) receptor kinase (Clark, Williams, & Meyerowitz 1997), a member of a large gene family (Lease, Ingham, & Walker 1998). Other members are *ERECTA* (*ER*) (Torii et al. 1996), *BRASSINOLIDE INSENSITVE* (*BRI*) (Li et al. 1997) and *HAESA* (*HAE*) (Jinn et al. 2000), genes which have a function in plant development

We investigated whether root specific *CLV1* homologs exist using a differential RT-PCR based approach. PCR products were generated from root tip and hypocotyl/cotyledon cDNA pools using degenerate primers, resulting in the isolation of the root specific *ROOT CLAVATA HOMOLOG 1* (*RCH1*). The promoter region of this gene was translationally fused to β-glucoronidase-green fluorescent protein (*GUS/GFP*) conferring root specific expression of these reporter genes (see Example 2).

Designing degenerate primers

To RT-PCR root specific members of the LRR receptor kinase gene family we compared the sequences of the *CLV1*, *ER* and *HAE* genes and their encoded proteins. The degenerate LRR primers were designed against the NxLxGxIP encoding region of the xLxxNxLxGxIPxxLxxLxxLxxL consensus for the LRR receptor-like kinases. Degenerate kinase primers were disigned for different homologous regions of the kinase domain. The primers are listed in Figure 2.

RNA Isolation

For each RNA isolation the roottips and the hypocotyl/cotyledon part including the shoot apical meristem of *Arabidopsis thaliana* (Col0) plants, 4 days after germination, were collected.

Total RNA was isolated (Pawlowski et al. 1994) and chromosomal DNA contamination was removed upon Dnase I (Promega, Madison, WI) treatment, following the instructions of the suppliers. The amount and quality of RNA was determined using spectrophotometer and agarose gel electrophoresis (data not shown).

Differential RT-PCR

Reverse transcription (RT) was performed using 5 µg of root tip (R) and hypocotyl/cotyledon (C) total RNA in a 20 µl reaction volume with 0.5 µg oligodT_{12-18} (Pharmacia, New Jersey, USA) and 200 U Superscript II (LifeTechnologies, Paisley, UK), according to the instructions of the supplier. Following first strand cDNA synthesis the cDNA samples were treated with RNase H (LifeTechnologies) as instructed by the supplier and subsequently diluted to a total volume of 100 µl with water.

Amplification reactions were performed with all degenerate LRR and KIN primer pairs (Figure 2). For the polymerase chain reaction (PCR) 2.5 µl of the R or C cDNA sample was amplified in a total volume of 50 µl reaction volume containing 100 µM dNTPs, 100 ng LRR and KIN primer, 1 U Taq polymerase (Boehringer, Ingelheim, Germany) and its accompanying buffer. The cDNAs were amplified during 40 cycles (1 min at 94°C, 1 min at 45-55°C, 3 min at 72°C).

Cloning and sequencing

The amplified R and C cDNAs were separated and compared with agarose gel electrophoresis (Figure 3). The root specific fragments, e.g. fragments present in R_{22} and not in C_{22}, were isolated from gel, cloned into the pGEM-T vector (Promega) and transformed into E. coli (strain DH5α), using standard cloning procedures.

Several PCR fragments of the same size but from different members of the LRR receptor kinase gene family may be generated using degenerate primer pairs. Therefore, 24 randomly picked colonies obtained after transformation were used in a "colony PCR". Aliquots of the amplified inserts were separated on agarose gel followed by Southern transfer to a Supercharge N+ membrane (Schleicher & Schuell, Dassel, Germany) to create a "colony blot". An aliquot of one insert was labeled with ^{32}P-dCTP and hybridized to the colony blot to determine the number of clones it represented in the pool of 24. This procedure was repeated with other labeled inserts until all clones were identified.

A single representative of each pool of clones was sequenced and the sequence was analyzed using the BLAST program. One of the clones sequenced was named *RCH1* and showed strong homology to plant LRR receptor kinases.

RT-PCR

The relative expression levels of *RCH1* in root tip (R), hypocotyl/cotyledon (C) and flowers (F) were determined using RT-PCR. Primers specific for the *RCH1* sequence were designed (RCH1F and RCH1R, Figure 2). The PCR samples were analyzed with agarose gel electrophoresis. *Ubiquitin* (*Ubi*) mRNA was used as an internal control (Horvath et al. 1993). Polyubiquitin genes consist of multiple units (Callis et al. 1995) and the *Ubi* primers hybridize with the ends of each single unit.

Independently isolated total RNA was reverse transcribed to obtain cDNA and for the PCR reaction 2 µl of the cDNA sample was amplified in a total volume of 50 µl as described abovefor the differential RT-PCR. The *RCH1* and *Ubi* cDNAs were amplified during 26 or 20 cycles (30 sec at 94°C, 30 sec at 56°C, 30 sec at 72°C), respectively (Figure 4). *RCH1* expression was only observed in root tip (R).

Example 2: Promoter isolation and promoter fusion construct.

To determine if the *RCH1* promoter region is sufficient to confer root specific expression

of a "gene of interest" we decided to fuse this promoter to the Glucuronidase/Green Fluorescent Protein (*GUS/GFP)* reporter gene. The *RCH1* clone was used to isolate a phage containing the corresponding genomic region from a λGEM (Promega, Madison, WI) based genomic library. The λ-GEM11 phage containing an insert of about 14550 nucleotides comprising the *Arabidopsis thaliana* RCH1 promoter and RCH1 gene and also the surrounding genomic sequences was deposited on October 25, 2000 at the BCCM-LMBP plasmid collection and was named "lambda phage RCH1 genomic". This deposit was given the accession number LMBP 5582CB by the international depositary authority. A restriction map was constructed of the genomic clone and a 7 kb XhoI fragment, containing 3.5 kb promoter and part of the *RCH1* coding sequence, was subcloned into the expression vector pBS (Stratagene, La Jolla, CA). The 3.5 kb promoter region including the first amino acids of the putative *RCH1* ORF is set forth in Figure 5 and is identified as SEQ ID NO 2. This promoter region was amplified using the primers M13F and PRCH1R and Pfu Taq polymerase (Stratagene, La Jolla, CA) in the manufacturer specified conditions. The promoter region was subsequently cloned into the binary pCAMBIA3383Xb vector (Roberts et al. 2000) creating a translational fusion with the *GUS/GFP* reporter gene.

GUS histochemical assay.

Arabidopsis thaliana plants were transformed with this construct according to the vacuum transformation method (Bechtold & Pelletier 1998). Transformants were selected for phosphinothricin-tripeptide (PTT, Duchefa, NL) resistance.

Histochemical localization of β -Glucuronidase activity (GUS) was performed using the substrate 5-bromo-4-chloro-3-indolyl glucuronide (X-gluc, Biosynth AG, Staad, Switzerland). X-gluc was dissolved in n-dimethyl-formamide and diluted to 0.5 mg/ml in 50 mM NaPO4, pH 7.2 with 0.1% Triton X-100. Oxidative dimerization of the produced indoxyl derivative was enhanced by adding the oxidation catalyst K^+ ferricyanide/ferrocyanide to a final concentration of 0.5 mM (0.5 mM $K_4Fe(CN)_6.H_2O$/0.5 mM $K_3Fe(CN)_6$). The GUS reaction was incubated at 37°C over several time periods, and pictures were taken from the cloured plant-tissues (Figures 7 and 15). In Figure 7 the very specific expression pattern of the RCH1 promoter is visualized by the GUS reporter gene activity. Persons skilled in the art will also agree to the fact that the expression level of the RCH 1 promoter is very high, since after one hour the saturation

Example 5: **Root-specific reporter gene expression in Rice mediated by the *Arabidopsis thaliana* RCH 1 promoter.**

CLAVATA homologs are found in many other plant species from both the groups of monocots and dicots. Therefore one can assume that also the promoter of RCH1 is functional in those other plant species, and possibly exerts the same specific expression-pattern. To investigate the feasibility of root meristem-specific activation of the *Arabidopsis thaliana* RCH1 promoter in monocots, the aforementioned promoter-reporter gene construct, wherein the GUS/GFP chimeric gene is operably linked to the RCH1 promoter, will be transformed to rice using the standard transformation procedures well known to the persons skilled in the art and outlined in the following paragraph. After several time periods ranging from 1 day to 1 or more weeks, the seedling will be checked for the expression of the GUS reporter gene. This will be done by growing the seedlings in organogenesis medium, and staining them with glucuronidase for several time periods ranging from 1 hour to 5 hours.

***Agrobacterium*-mediated rice transformation**

The Promoter-GUS hybrid gene outlined in Example 2 can be transformed to *Agrobacterium tumefaciens* strain LBA4404 or C58 by means of electroporation and subsequently transformed bacterial cells can be selected on a solid agar medium containing the appropriate antibiotics.

For demonstration of root-specific GUS expression with the promoter of the current invention, 309 mature dry seeds of the rice japonica cultivars Nipponbare or Taipei are dehusked, sterilised and germinated on a medium containing 2,4-dichlorophenoxyacetic acid (2,4-D). After incubation in the dark for four weeks, embryogenic, scutellum-derived calli are excised and propagated on the same medium. Selected embryogenic callus is then co-cultivated with *Agrobacterium*. Co-cultivated callus is grown on 2,4-D-containing medium for 4 to 5 weeks in the dark in the presence of a suitable concentration of the appropriate selective agent. During this period, rapidly growing resistant callus islands develop. After transfer of this material to a medium with a reduced concentration of 2,4-D and incubation in the light, the embryogenic potential is released and shoots develop in the next four to five weeks. Shoots are excised from the callus and incubated for one week on an auxin-containing medium from which they can be transferred to the soil. Hardened shoots are grown under high humidity and short days in a phytotron. Seeds

level of staining is reached in the meristem of the main root from a 1 week old seedling (Figure 7 D).

RCH1::GFP expression in roots is shown in Figure 16.

Example 3: Alignment.

Sequence alignment of one complete register sequence against a target sequence as in figure 10 was done using the program GAP of the GCG package. The algorithm of Needleman and Wunsch is applied here to find the alignment of two complete sequences (Needleman and Wunsch, JMB 48(3): 443-453, 1970). The used parameters during the alignment were Gap Weight: 8, Average Match: 2.912, Length Weight: 2, Average Mismatch: -2.003, Quality: 1388, Length: 1149, Ratio: 1.416, and Gaps: 25.

Example 4: Phylogenetic tree.

For sequence comparison of proteins as in figure 11 and 12, the Clustal V method to align all the sequences was used (Higgins DG & Sharp PM, *Comput Appl Biosci.* 1989 Apr;5(2):151-3). The Clustal Method groups sequences into clusters by examining sequence distances between all pairs. Clusters are aligned as pairs then collectively as sequence groups to produce the overall alignment. After the multiple alignment is completed, a neighbor-joining method is employed to reconstruct phylogeny for the putative alignment.

The Clustal method uses weight tables to construct multiple alignments. Residue weight tables are used in scoring protein and nucleotide alignments so that slightly mismatched residues, such as Ile vs. Val amino acids or G vs. R nucleotides, are not scored the same as total mismatches. For the alignment and pylogenetic trees in figure 11 and 12 the PAM250 weight table was used. Other method parameters for pairwise alignment were K-tuple: 1, Gap Penalty: 3, Window: 5 and Diagonals: 5. Other method parameters for multiple alignment include Gap Penalty: 10 and Gap Length Penalty: 10.

To view phylogenetic relationships compatible with multiple sequence alignments a Phylogenetic Tree was calculated. The length of each pair of branches represents the distance between sequence pairs, while the units at the bottom of the tree indicate the number of substitution events. The phylogenetic tree gives a first estimate of the relationship between homologous sequences. A cladogram presentation of the phylogenetic tree is shown in figures 11 and 12.

can be harvested three to five months after transplanting. The method yields single locus transformants at a rate of over 50 % (Chan et al. 1993;Hiei et al. 1994)

Example 6: Expression of a cytokinine oxidase cDNA under the root-specific RCH1 promoter in transgenic Tobacco and Rice leads to increased root production.

The AtCKX1 gene (located in the database with the accession number: AC002510, *Arabidosis thaliana* chromosome II section 225 of 255 of the complete sequence, sequence from clone T32G6) can be cloned under the control of the RCH1 promoter of Arabidopsis, in a binary vector. Also the ZmCKX1 from maize (accession number AF044603, (Morris et al. 1999) can be operably linked to the RCH1 promoter a binary vector. These genes can be transfected into Tobacco or Rice plants, using agrobacterium-mediated gene transfer. The transgenic plants expressing the ATCKX1 or the ZmCKX1 specifically in roots are expected to show increased root production without negatively affecting shoot development and without inducing premature leaf senescence.

Example 7. Minimal promoter sequence.

The length of the RCH1 promoter region (SEQ ID NO 2), which confers to root specific expression patterns, can be reduced to a "minimal promoter" without altering the root specific expression patterns or other features of the promoter such as constitutive and strong activity. As an example, the length of the promoter region was reduced to a minimal promoter sequence ranging from the –1 base pair just before the start codon at position 3479-3481 of SEQ ID NO. 2 to -842 base pairs before said start codon. Said minimal promoter is represented in SEQ ID NO. 18. This minimal promoter was PCR amplified and translationally fused to the GUS reporter gene. With a GUS histochemical assay (example 2) the same expression pattern as the large promoter region was demonstrated (results not shown). The difference is that the sequence before the HindIII site at the beginning of the minimal promoter is not necessary for the root specific expression pattern.

Example 8. RCH1 prom::GUS expression in embryo's.

Siliques of Arabidopsis plants transformed with RCH1prom::GUS were opened at different stages of development and the embryo's were surgically taken out of the ovules and transferred to X-gluc staining solution. Embryo's were stained for 0.5h – 4h before pictures were taken using light microscopy, Fig 15 A and B are at 25X and 40X magnification respectively, stained for 0.5 h. Longer staining did not reveal expression outside the root promeristem region.

RCH1 is active very early: RCH1 expression was first observed in torpedo stage embryo's in the root promeristem and was maintained in this region during the later stages of development. This figure shows that as soon as a meristem is formed RCH1 is expressed and provides further evidence that RCH1 expression is correlated with the appearance and presence of a root meristem from the embryonic stage onwards.

Example 9. Isolation of cDNA of other Arabidopsis RCH1 homologues

Table 4: % identity with RCH1

	Database entry genomic DNA	Database entry protein	% sequence identity with RCH1 on aa level
RCH1.2	AB028621 + AP002037	BAB03091	59.8%
RCH1.3	AB011476	BAB09286	40.1%
RCH1.4	AL022223	CAA1826 or T05050	38.9%
RCH1. 5	AC015446	AAG12526	42.1%

To discover the exact function of RCH1, the inventors want to circumvent the redundancy of the gene.

AB028621 (see figure 12, phylogenetic tree) is determined to be RCH1.2 protein and is also used by the inventors for further experiments. Double knock-outs of RCH1 and RCH1.2 were made and also other closely related sequences (AB011476.pro,

T05050.pro and AC015446.pro) are used in knock out experiments or in antisense expression.

Ectopic expression of RCH1

The expression of RCH1 in normal cells correlates with the zone of rapidly dividing cells and is involved in the maintenance of meristem cells and dividing cells. Improved activity of RCH1 in a host cell results in the maintenance of cell division in that host cell and thus works against the differentiation of that cell. On the contrary diminished RCH1 activity in a host cell results in a diminished cell division in that host cell and thus in the loss of the amount of dividing cells in the meristem. Particular examples to influence these processes by altering the expression level of the RCH1 gene are described in the following paragraphs:

Overexpression

In order to obtain higher activity of the RCH1 protein in the host cell the coding region of the RCH1 gene is fused to a strong promoters that is inducible or constitutive and/or ubiquitous or tissue specific.

To influence the RCH1 activity in a particular set of host cells, the RCH1 coding region is cloned under control of other meristemic promoters such as the promoters of WUSCHEL, CLV1, CLV2, CLV3, STM. The obtained effect is the control of the expression of RCH1 in a particular area of the meristem or in a particular time period of the development of the plant. Also the promoters of other RCH genes (such as AB028621.pro, AB011476.pro, T05050.pro and AC015446.pro) are used for this purpose.

In a particular application of the present invention, the RCH1 gene and the RKN gene (= the rice homologue of RCH1 as described in WO/04761) are operationally linked to the root-specific promoter PYK10, a constitutive ubiquitin or GOS 2 promoter and Cdc2a promoter and expressed in Arabidopsis or in rice.

Downregulation

In order to obtain lower activity of the RCH1 protein in the host cell, the coding region of the RCH1 gene cloned in a sense and antisense orientation in order to form a hairpin that induces gene-silencing in the host cell.

Co-expression

Downregulation or upregulation of the RCH1 in combination with downregulation or upregulation of other genes e.g.:

With WUSHEL:

In the shoot meristem a balance between the *WUSCHEL* and *CLAVATA* (*CLV*) genes has been implicated to play an essential role in regulation cell division and differentiation (Clark et al. 1993;Fletcher et al. 1999;Schoof et al. 2000). The whole meristem maintenance is controlled by these genes and they seem to have an opposite function keeping the balance between cell division and differentiation. Therefore, WUSCHEL or a root WUSCHEL homologue is a preferred partner to be co-expressed with RCH1.

In an application of the invention RCH1 is upregulated in combination with the downregulation of WUSCHEL in a root specific manner. The obtained effect of this co-expression is a better control on the meristem maintenance.

With other RCH1 family members

Possibly the function of RCH1 is redundant and other closely related genes can take over its function. Therefore the RCH1 gene ectopically expressed in combination with closely related genes in order to downregulate or upregulate the activity of the encoded protein. The RCH1 homologues used for co-expression are for example AB028621.pro, AB011476.pro, T05050.pro and AC015446 (see figure 12 and table 4). The obtained effect is a better effect as described for the ectopic expression of RCH1 alone.

With other Clavata homologues:

Clavata 1 is a receptor that is involved in a signal pathway probably through the binding of a ligand (e.g. Clavata 3). A particular useful application of the present invention is the co-expression of RCH1 with the root Clavata 3 homologue RCH3. The obtained effect is a better control on meristem maintenance.

Example 10: Domain swapping

The N-terminal LRR domain of RCH1 functions as an extracellular perception domain and a C-terminal kinase domain functions in an intracellular signal transduction pathway. Previously it was described by He et al. (Science 2000, 288(5475):2360-3: Perception of brassinosteroids by the extracellular domain of the receptor kinase BRI1) that such domains can be swapped with the domains of other proteins. Following this methodology in a particular application of the present invention, the domains of RCH1 are swapped with the domains of another protein. In one example the LRR domain of RCH1 is replaced by the LRR domain of BRI. The effect obtained is the activation of a different pathway, which is normally not activated by RCH1, because the kinase domain of RCH1 now responds to another external signal that normally is not recognized by RCH1.

It will be clear that the invention may be practiced otherwise than as particularly described in the forgoing description and examples. Numerous modifications and variations of the present invention are possible in light of the above teaching and, therefore, are within the scope of the appended claims.

REFERENCES

Alberts,B. et al. (1994) Molecular Biology of the Cell, Garland Publishing Inc..

Altschul S.F. et al. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389-3402.

An G. et al. (1985) New cloning vehicles for transformation of higher plants. *EMBO J.* **4**, 277-284.

Armstrong C.L. et al. (1990) Factors affecting PEG-mediated stable transformation of maize protoplasts. *Plant Cell Reports* **9**, 335-339.

Banerjee A. et al. (1996) Omega amino acids in peptide design: incorporation into helices. *Biopolymers* **39**, 769-777.

Baron M.H. & Baltimore D. (1982) Antibodies against the chemically synthesized genome-linked protein of poliovirus react with native virus-specific proteins. *Cell* **28**, 395-404.

Bartel,P.L. & Fields,S. (1997) The Yeast Two-Hybrid System, Oxford University Press.

Baszczynski, C. L., Fallis, L., Bellemare, G., and Boivin, R. Brassica regulatory sequence for root-specific or root-abundant gene expression. WO9007001. 2000.

Bechtold N. & Pelletier G. (1998) In planta Agrobacterium-mediated transformation of adult Arabidopsis thaliana plants by vacuum infiltration. *Methods Mol.Biol.* **82**, 259-266.

Bell M.H. et al. (1993) Tobacco plants transformed with cdc25, a mitotic inducer gene from fission yeast. *Plant Mol.Biol* **23**, 445-451.

Benkirane N. et al. (1996) Exploration of requirements for peptidomimetic immune recognition. Antigenic and immunogenic properties of reduced peptide bond pseudopeptide analogues of a histone hexapeptide. *J.Biol Chem.* **271**, 33218-33224.

Berry A. & Brenner S.E. (1994) A prototype computer system for de novo protein design. *Biochem.Soc.Trans.* **22**, 1033-1036.

Binarova P. et al. (1998) Treatment of Vicia faba root tip cells with specific inhibitors to cyclin-dependent kinases leads to abnormal spindle formation. *Plant J.* **16**, 697-707.

Bögre L. et al. (1997) The cdc2Ms kinase is differentially regulated in the cytoplasm and in the nucleus. *Plant Physiol.* **113**, 841-852.

Bögre L. et al. (1999) A MAP kinase is activated late in plant mitosis and becomes localized to the plane of cell division. *Plant Cell* **11**, 101-113.

Borg S. et al. (1997) Identification of new protein species among 33 different small GTP-binding proteins encoded by cDNAs from Lotus japonicus, and expression of corresponding mRNAs in developing root nodules. *Plant J.* **11**, 237-250.

Calderini O. et al. (1998) A cell cycle regulated MAP kinase with a possible role in cytokinesis in tobacco cells. *J.Cell Sci.* **111 (Pt 20)**, 3091-3100.

Callis J. et al. (1995) Structure and evolution of genes encoding polyubiquitin and ubiquitin- like proteins in Arabidopsis thaliana ecotype Columbia. *Genetics* **139**, 921-939.

Chan M.T. et al. (1993) Agrobacterium-mediated production of transgenic rice plants expressing a chimeric alpha-amylase promoter/beta-glucuronidase gene. *Plant Mol.Biol.* **22**, 491-506.

Chaubet N. et al. (1996) Identification of cis-elements regulating the expression of an Arabidopsis histone H4 gene. *Plant J.* **10**, 425-435.

Christou P., McCabe D.E., & Swain W.F. (1988) Stable transformation of soybean callus by DNA-coated gold particles. *Plant Physiol.* **87**, 671-674.

Clark S.E., Running M.P., & Meyerowitz E.M. (1993) CLAVATA1, a regulator of meristem and flower development in Arabidopsis. *Development* **119**, 397-418.

Clark S.E., Williams R.W., & Meyerowitz E.M. (1997) The CLAVATA1 gene encodes a putative receptor kinase that controls shoot and floral meristem size in Arabidopsis. *Cell* **89**, 575-585.

Cohen-Fix O. & Koshland D. (1997) The metaphase-to-anaphase transition: avoiding a mid-life crisis. *Curr.Opin.Cell Biol* **9**, 800-806.

Colasanti J., Tyers M., & Sundaresan V. (1991) Isolation and characterization of cDNA clones encoding a functional p34cdc2 homologue from Zea mays. *Proc.Natl.Acad.Sci.U.S.A* **88**, 3377-3381.

Crossway A. et al. (1986) Integration of foreign DNA following microinjection of tobacco mesophyll protoplasts. *Mol.Gen.Genet.* **202**, 179-185.

Dale E.C. & Ow D.W. (1990) Intra- and intermolecular site-specific recombination in plant cells mediated by bacteriophage P1 recombinase. *Gene* **91**, 79-85.

de Pater B.S. & Schilperoort R.A. (1992) Structure and expression of a root-specific rice gene. *Plant Mol.Biol.* **18**, 161-164.

De Veylder L. et al. (1997) The Arabidopsis Cks1At protein binds the cyclin-dependent kinases Cdc2aAt and Cdc2bAt. *FEBS Lett.* **412**, 446-452.

Dodds,J.H. (1985) Plant genetic engineering, Cambridge University Press.

Doerner P. et al. (1996) Control of root growth and development by cyclin expression. *Nature* **380**, 520-523.

Dolferus R. et al. (1994) Differential interactions of promoter elements in stress responses of the Arabidopsis Adh gene. *Plant Physiol* **105**, 1075-1087.

Dorner B. et al. (1996) The synthesis of peptidomimetic combinatorial libraries through successive amide alkylations. *Bioorg.Med.Chem.* **4**, 709-715.

Elledge S.J. (1996) Cell cycle checkpoints: preventing an identity crisis. *Science* **274**, 1664-1672.

Ellis J.G. et al. (1987) Maize Adh-1 promoter sequences control anaerobic regulation: addition of upstream promoter elements from constitutive genes is necessary for expression in tobacco. *EMBO J.* **6**, 11-16.

Ely, S., Evans, I. J., and Schuch, W. W. Root-specific promotor. 288630. 2000. 8-10-1994. Ref Type: Patent

Evans T. et al. (1983) Cyclin: a protein specified by maternal mRNA in sea urchin eggs that is destroyed at each cleavage division. *Cell* **33**, 389-396.

Fantes P. (1989) Yeast cell cycle. *Curr.Opin.Cell Biol* **1**, 250-255.

Fassina G. & Melli M. (1994) Identification of interactive sites of proteins and protein receptors by computer-assisted searches for complementary peptide sequences. *Immunomethods.* **5**, 114-120.

Fedoroff N.V. & Smith D.L. (1993) A versatile system for detecting transposition in Arabidopsis. *Plant J.* **3**, 273-289.

Feiler H.S. & Jacobs T.W. (1990) Cell division in higher plants: a cdc2 gene, its 34-kDa product, and histone H1 kinase activity in pea. *Proc.Natl.Acad.Sci.U.S.A* **87**, 5397-5401.

Fesquet D. et al. (1993) The MO15 gene encodes the catalytic subunit of a protein kinase that activates cdc2 and other cyclin-dependent kinases (CDKs) through phosphorylation of Thr161 and its homologues. *EMBO J.* **12**, 3111-3121.

Fletcher J.C. et al. (1999) Signaling of cell fate decisions by CLAVATA3 in Arabidopsis shoot meristems. *Science* **283**, 1911-1914.

Fowler T.J., Bernhardt C., & Tierney M.L. (1999) Characterization and expression of four proline-rich cell wall protein genes in Arabidopsis encoding two distinct subsets of multiple domain proteins. *Plant Physiol* **121**, 1081-1092.

Francis D. & Halford N.G. (1995) The plant cell cycle. *Physiol.Plant.* **93**, 365-374.

Francis,D., Dudits,D., & Inzé,D. (1998) Plant Cell Division, Portland Press London.

Fromm M., Taylor L.P., & Walbot V. (1985) Expression of genes transferred into monocot and dicot plant cells by electroporation. *Proc.Natl.Acad.Sci.U.S.A* **82**, 5824-5828.

Giritch A. et al. (1997) A root-specific iron-regulated gene of tomato encodes a lysyl-tRNA- synthetase-like protein. *Eur.J.Biochem.* **244**, 310-317.

Goddemeier M.L., Wulff D., & Feix G. (1998) Root-specific expression of a Zea mays gene encoding a novel glycine- rich protein, zmGRP3. *Plant Mol.Biol.* **36**, 799-802.

Gorthy P. & Pudukottah M.G. (1999) Production of silicon carbide from rice husks. *J.Am.Ceramic Soc.* **82**.

Gubler F. & Jacobsen J.V. (1992) Gibberellin-responsive elements in the promoter of a barley high-pI alpha-amylase gene. *Plant Cell* **4**, 1435-1441.

Hansen G. & Chilton M.D. (1996) "Agrolistic" transformation of plant cells: integration of T-strands generated in planta. *Proc.Natl.Acad.Sci.U.S.A* **93**, 14978-14983.

Hansen G., Shillito R.D., & Chilton M.D. (1997) T-strand integration in maize protoplasts after codelivery of a T-DNA substrate and virulence genes. *Proc.Natl.Acad.Sci.U.S.A* **94**, 11726-11730.

Hansen G., Shillito R.D., & Chilton M.D. (1997) T-strand integration in maize protoplasts after codelivery of a T-DNA substrate and virulence genes. *Proc.Natl.Acad.Sci.U.S.A* **94**, 11726-11730.

Hanson B. et al. (1999) A simple method to enrich an Agrobacterium-transformed population for plants containing only T-DNA sequences. *Plant J.* **19**, 727-734.

Harlow,E. & Lane,D. (1988) Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory Press.

Haseloff J. et al. (1997) Removal of a cryptic intron and subcellular localization of green fluorescent protein are required to mark transgenic Arabidopsis plants brightly. *Proc.Natl.Acad.Sci.U.S.A* **94**, 2122-2127.

Hayles J. & Nurse P. (1986) Cell cycle regulation in yeast. *J.Cell Sci.Suppl* **4**, 155-170.

Helariutta Y. et al. (2000) The SHORT-ROOT gene controls radial patterning of the Arabidopsis root through radial signaling [In Process Citation]. *Cell* **101**, 555-567.

Hemerly A.S. et al. (1993) cdc2a expression in Arabidopsis is linked with competence for cell division. *Plant Cell* **5**, 1711-1723.

Herrera-Estrella L. et al. (1983) Chimeric genes as dominant selectable markers in plant cells. *EMBO J.* **2**, 987-995.

Herrera-Estrella L. et al. (1983) Expression of chimaeric genes transferred into plant cells using a Ti-plasmid-derived vector. *Nature* **303**, 209-213.

Hiei Y. et al. (1994) Efficient transformation of rice (Oryza sativa L.) mediated by Agrobacterium and sequence analysis of the boundaries of the T-DNA. *Plant J.* **6**, 271-282.

Hirt H. et al. (1991) Complementation of a yeast cell cycle mutant by an alfalfa cDNA encoding a protein kinase homologous to p34cdc2. *Proc.Natl.Acad.Sci.U.S.A* **88**, 1636-1640.

Hirt H. et al. (1993) cdc2MsB, a cognate cdc2 gene from alfalfa, complements the G1/S but not the G2/M transition of budding yeast cdc28 mutants. *Plant J.* **4**, 61-69.

Hochstrasser M. (1998) There's the rub: a novel ubiquitin-like modification linked to cell cycle regulation. *Genes Dev.* **12**, 901-907.

Hoffman D.L. et al. (1995) Rapid protein structure classification using one-dimensional structure profiles on the bioSCAN parallel computer. *Comput.Appl.Biosci.* **11**, 675-679.

Horvath B. et al. (1993) Lipo-oligosaccharides of Rhizobium induce infection-related early nodulin gene expression in pea root hairs. *Plant J.* **4**, 727-733.

Huntley R. et al. (1998) The maize retinoblastoma protein homologue ZmRb-1 is regulated during leaf development and displays conserved interactions with G1/S regulators and plant cyclin D (CycD) proteins. *Plant Mol.Biol.* **37**, 155-169.

Hwang I. & Goodman H.M. (1995) An Arabidopsis thaliana root-specific kinase homolog is induced by dehydration, ABA, and NaCl. *Plant J.* **8**, 37-43.

Jaglo-Ottosen K.R. et al. (1998) Arabidopsis CBF1 overexpression induces COR genes and enhances freezing tolerance. *Science* **280**, 104-106.

Jeong S., Trotochaud A.E., & Clark S.E. (1999) The Arabidopsis CLAVATA2 gene encodes a receptor-like protein required for the stability of the CLAVATA1 receptor-like kinase. *Plant Cell* **11**, 1925-1934.

Jinn T.L., Stone J.M., & Walker J.C. (2000) HAESA, an Arabidopsis leucine-rich repeat receptor kinase, controls floral organ abscission. *Genes Dev.* **14**, 108-117.

John P.C., Sek F.J., & Lee M.G. (1989) A homolog of the cell cycle control protein p34cdc2 participates in the division cycle of Chlamydomonas, and a similar protein is detectable in higher plants and remote taxa. *Plant Cell* **1**, 1185-1193.

John,P.C.L. (1981) The cell cycle, Cambridge University Press.

Joung J.K., Ramm E.I., & Pabo C.O. (2000) A bacterial two-hybrid selection system for studying protein-DNA and protein-protein interactions. *Proc.Natl.Acad.Sci.U.S.A* **97**, 7382-7387.

Keller B. & Lamb C.J. (1989) Specific expression of a novel cell wall hydroxyproline-rich glycoprotein gene in lateral root initiation. *Genes Dev.* **3**, 1639-1646.

Kikuchi H. et al. (1999) Molecular characterization of a gene for alanine aminotransferase from rice (Oryza sativa). *Plant Mol.Biol.* **39**, 149-159.

Krek W. (1998) Proteolysis and the G1-S transition: the SCF connection. *Curr.Opin.Genet.Dev.* **8**, 36-42.

Krens F.A. et al. (1982) *In vitro* transformation of plant protoplasts with Ti-plasmid DNA. *Nature* **296**, 72-74.

Kumagai A. & Dunphy W.G. (1991) The cdc25 protein controls tyrosine dephosphorylation of the cdc2 protein in a cell-free system. *Cell* **64**, 903-914.

Labbe J.C. et al. (1989) MPF from starfish oocytes at first meiotic metaphase is a heterodimer containing one molecule of cdc2 and one molecule of cyclin B. *EMBO J.* **8**, 3053-3058.

Lagarde D. et al. (1996) Tissue-specific expression of Arabidopsis AKT1 gene is consistent with a role in K+ nutrition. *Plant J.* **9**, 195-203.

Lake R.S. & Salzman N.P. (1972) Occurrence and properties of a chromatin-associated F1-histone phosphokinase in mitotic Chinese hamster cells. *Biochemistry* **11**, 4817-4826.

Langan T.A. (1978) Methods for the assessment of site-specific histone phosphorylation. *Methods Cell Biol* **19**, 127-142.

Lauter F.R. (1996) Root-specific expression of the LeRse-1 gene in tomato is induced by exposure of the shoot to light. *Mol.Gen.Genet.* **252**, 751-754.

Lease K., Ingham E., & Walker J.C. (1998) Challenges in understanding RLK function. *Curr.Opin.Plant Biol.* **1**, 388-392.

Lee M.G. & Nurse P. (1987) Complementation used to clone a human homologue of the fission yeast cell cycle control gene cdc2. *Nature* **327**, 31-35.

Lerner R.A. (1982) Tapping the immunological repertoire to produce antibodies of predetermined specificity. *Nature* **299**, 593-596.

Lerner R.A. et al. (1981) Chemically synthesized peptides predicted from the nucleotide sequence of the hepatitis B virus genome elicit antibodies reactive with the native envelope protein of Dane particles. *Proc.Natl.Acad.Sci.U.S.A* **78**, 3403-3407.

Li J. & Chory J. (1997) A putative leucine-rich repeat receptor kinase involved in brassinosteroid signal transduction. *Cell* **90**, 929-938.

Liddle,J.E. & Cryer,A. (1991) A Practical Guide to Monoclonal Antibodies, Wiley New York.

Lisztwan J. et al. (1998) Association of human CUL-1 and ubiquitin-conjugating enzyme CDC34 with the F-box protein p45(SKP2): evidence for evolutionary conservation in the subunit composition of the CDC34-SCF pathway. *EMBO J.* **17**, 368-383.

Llewellyn D.J. et al. (1987) Structure and expression of an alcohol dehydrogenase 1 gene from Pisum sativum (cv. "Greenfeast"). *J.Mol.Biol.* **195**, 115-123.

Loffer F.D. (1994) The need to monitor intrauterine pressure--myth or necessity? [editorial; comment]. *J.Am.Assoc.Gynecol.Laparosc.* **2**, 1-2.

Long J.A. & Barton M.K. (1998) The development of apical embryonic pattern in Arabidopsis. *Development* **125**, 3027-3035.

Lundgren K. et al. (1991) mik1 and wee1 cooperate in the inhibitory tyrosine phosphorylation of cdc2. *Cell* **64**, 1111-1122.

Magyar A. & Mihalik R. (1997) Characterisation of the swine swC1 antigen. *Acta Vet.Hung.* **45**, 17-31.

Malamy J.E. & Benfey P.N. (1997) Analysis of SCARECROW expression using a rapid system for assessing transgene expression in Arabidopsis roots. *Plant J.* **12**, 957-963.

Masuda S., Sakuta C., & Satoh S. (1999) cDNA cloning of a novel lectin-like xylem sap protein and its root- specific expression in cucumber. *Plant Cell Physiol* **40**, 1177-1181.

Merrifield R.B. (1963) Solid phase peptide synthesis. I. The synthesis of a tetrapeptide. *J.Amer.Chem.Soc.* **85**, 2149-2154.

Mironov V. et al. (1999) Cyclin-dependent kinases and cell division in plants- the nexus. *Plant Cell* **11**, 509-522.

Monge A. et al. (1995) Computer modeling of protein folding: conformational and energetic analysis of reduced and detailed protein models. *J.Mol.Biol* **247**, 995-1012.

Morris R.O. et al. (1999) Isolation of a gene encoding a glycosylated cytokinin oxidase from maize. *Biochem.Biophys.Res.Commun.* **255**, 328-333.

Murakami T., Simonds W.F., & Spiegel A.M. (1992) Site-specific antibodies directed against G protein beta and gamma subunits: effects on alpha and beta gamma subunit interaction. *Biochemistry* **31**, 2905-2911.

Murray A.W. & Kirschner M.W. (1989) Dominoes and clocks: the union of two views of the cell cycle. *Science* **246**, 614-621.

Nasmyth K. (1993) Control of the yeast cell cycle by the Cdc28 protein kinase. *Curr.Opin.Cell Biol* **5**, 166-179.

Norbury C. & Nurse P. (1992) Animal cell cycles and their control. *Annu.Rev.Biochem.* **61**, 441-470.

Nurse P. & Bissett Y. (1981) Gene required in G1 for commitment to cell cycle and in G2 for control of mitosis in fission yeast. *Nature* **292**, 558-560.

Nurse P. (1990) Universal control mechanism regulating onset of M-phase. *Nature* **344**, 503-508.

Okada K. et al. (2000) Five geranylgeranyl diphosphate synthases expressed in different organs are localized into three subcellular compartments in arabidopsis [In Process Citation]. *Plant Physiol* **122**, 1045-1056.

Okumura N. et al. (1994) A dioxygenase gene (Ids2) expressed under iron deficiency conditions in the roots of Hordeum vulgare. *Plant Mol.Biol.* **25**, 705-719.

Olson D.C., Oetiker J.H., & Yang S.F. (1995) Analysis of LE-ACS3, a 1-aminocyclopropane-1-carboxylic acid synthase gene expressed during flooding in the roots of tomato plants. *J.Biol.Chem.* **270**, 14056-14061.

Olszewski K.A., Kolinski A., & Skolnick J. (1996) Folding simulations and computer redesign of protein A three-helix bundle motifs. *Proteins* **25**, 286-299.

Onouchi H. et al. (1991) Operation of an efficient site-specific recombination system of Zygosaccharomyces rouxii in tobacco cells. *Nucleic Acids Res.* **19**, 6373-6378.

Onouchi H. et al. (1995) Visualization of site-specific recombination catalyzed by a recombinase from Zygosaccharomyces rouxii in Arabidopsis thaliana. *Mol.Gen.Genet.* **247**, 653-660.

Ormrod,J.C. & Francis,D. (1993) Molecular and Cell Biology of the Plant Cell Cycle, Kluwer Academic Publishers.

Osborne B.I., Wirtz U., & Baker B. (1995) A system for insertional mutagenesis and chromosomal rearrangement using the Ds transposon and Cre-lox. *Plant J.* **7**, 687-701.

Ostresh J.M. et al. (1996) Generation and use of nonsupport-bound peptide and peptidomimetic combinatorial libraries. *Methods Enzymol.* **267**, 220-234.

Pabo C.O. & Suchanek E.G. (1986) Computer-aided model-building strategies for protein design. *Biochemistry* **25**, 5987-5991.

Pawlowski K. et al. (1994) Isolation of total, poly(A) and polysomal RNA from plant tissue. In: Plant Molecular Biology Manual (eds S.B.Gelvin & R.A.Schilperoort), pp. 1-13. Kluwer academic Publisher, dordrecht.

Peralta E.G., Hellmiss R., & Ream W. (1986) Overdrive, a T-DNA transmission enhancer on the A. tumefaciens tumour-inducing plasmid. *EMBO J.* **5**, 1137-1142.

Pines J. (1995) Cell cycle. Confirmational change [news; comment]. *Nature* **376**, 294-295.

Poon R.Y. et al. (1993) The cdc2-related protein p40MO15 is the catalytic subunit of a protein kinase that can activate p33cdk2 and p34cdc2. *EMBO J.* **12**, 3123-3132.

Pysh L.D. et al. (1999) The GRAS gene family in Arabidopsis: sequence characterization and basic expression analysis of the SCARECROW-LIKE genes. *Plant J.* **18**, 111-119.

Reed S.I., Hadwiger J.A., & Lorincz A.T. (1985) Protein kinase activity associated with the product of the yeast cell division cycle gene CDC28. *Proc.Natl.Acad.Sci.U.S.A* **82**, 4055-4059.

Renaudin J.P. et al. (1996) Plant cyclins: a unified nomenclature for plant A-, B- and D-type cyclins based on sequence organization. *Plant Mol.Biol.* **32**, 1003-1018.

Renouf D.V. & Hounsell E.F. (1995) Molecular modelling of glycoproteins by homology with non-glycosylated protein domains, computer simulated glycosylation and molecular dynamics. *Adv.Exp.Med.Biol* **376**, 37-45.

Roberts, C. S., Rajagopal, S., Smith, L. A., Nguyen, T. A., Yang, W., Nugroho, S., Ravi, K. S., Cao, M-L., Vijayachandra, K., Patell, V., Harcourt, R. L., Dransfield, L., Desamero, N., Slamet, I., Keese, P., Kilian, A., and Jefferson, R. A. A comprehensive set of modular vectors for advanced manipulations and efficient transformation of

plants by both *Agrobacterium* and direct DNA uptake methods. 2000.

Rose R.B. et al. (1996) Three-dimensional structures of HIV-1 and SIV protease product complexes. *Biochemistry* **35**, 12933-12944.

Russell P. & Nurse P. (1986) cdc25+ functions as an inducer in the mitotic control of fission yeast. *Cell* **45**, 145-153.

Russell P. & Nurse P. (1987) Negative regulation of mitosis by wee1+, a gene encoding a protein kinase homolog. *Cell* **49**, 559-567.

Russell P. & Nurse P. (1987) The mitotic inducer nim1+ functions in a regulatory network of protein kinase homologs controlling the initiation of mitosis. *Cell* **49**, 569-576.

Rutenber E.E. et al. (1996) A new class of HIV-1 protease inhibitor: the crystallographic structure, inhibition and chemical synthesis of an aminimide peptide isostere. *Bioorg.Med.Chem.* **4**, 1545-1558.

Sambrook,J., Fritsch,E.F., & Maniatis,T. (1989) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press.

Schlappi M., Smith D., & Fedoroff N. (1993) TnpA trans-activates methylated maize Suppressor-mutator transposable elements in transgenic tobacco. *Genetics* **133**, 1009-1021.

Schneider K. et al. (1998) The ROOT HAIRLESS 1 gene encodes a nuclear protein required for root hair initiation in Arabidopsis. *Genes Dev.* **12**, 2013-2021.

Schnittger A., Schellmann S., & Hulskamp M. (1999) Plant cells - young at heart? *Curr.Opin.Plant Biol.* **2**, 508-512.

Schoof H. et al. (2000) The stem cell population of Arabidopsis shoot meristems in maintained by a regulatory loop between the CLAVATA and WUSCHEL genes. *Cell* **100**, 635-644.

Semler B.L. et al. (1982) A membrane-associated precursor to poliovirus VPg identified by immunoprecipitation with antibodies directed against a synthetic heptapeptide. *Cell* **28**, 405-412.

Shioda T. et al. (2000) A green fluorescent protein-reporter mammalian two-hybrid system with extrachromosomal maintenance of a prey expression plasmid: Application to interaction screening. *Proc.Natl.Acad.Sci.U.S.A* **97**, 5220-5224.

Soni R. et al. (1995) A family of cyclin D homologs from plants differentially controlled by growth regulators and containing the conserved retinoblastoma protein interaction motif. *Plant Cell* **7**, 85-103.

Sorrell D.A. et al. (1999) Distinct cyclin D genes show mitotic accumulation or constant levels of transcripts in tobacco bright yellow-2 cells. *Plant Physiol* **119**, 343-352.

Sugita K. et al. (2000) Technical advance: A transformation vector for the production of marker- free transgenic plants containing a single copy transgene at high frequency [In Process Citation]. *Plant J.* **22**, 461-469.

Sun Y. et al. (1999) Characterization of maize (Zea mays L.) Wee1 and its activity in developing endosperm. *Proc.Natl.Acad.Sci.U.S.A* **96**, 4180-4185.

Suzuki H., Fowler T.J., & Tierney M.L. (1993) Deletion analysis and localization of SbPRP1, a soybean cell wall protein gene, in roots of transgenic tobacco and cowpea. *Plant Mol.Biol.* **21**, 109-119.

Swenson K.I., Farrell K.M., & Ruderman J.V. (1986) The clam embryo protein cyclin A induces entry into M phase and the resumption of meiosis in Xenopus oocytes. *Cell* **47**, 861-870.

Tamura R.N. et al. (1991) Cell type-specific integrin variants with alternative alpha chain cytoplasmic domains. *Proc.Natl.Acad.Sci.U.S.A* **88**, 10183-10187.

Torii K.U. et al. (1996) The Arabidopsis ERECTA gene encodes a putative receptor protein kinase with extracellular leucine-rich repeats. *Plant Cell* **8**, 735-746.

Trieu A.T. et al. (2000) Technical Advance: Transformation of Medicago truncatula via infiltration of seedlings or flowering plants with Agrobacterium. *Plant J.* **22**, 531-541.

Uribe X. et al. (1998) Maize alpha-tubulin genes are expressed according to specific patterns of cell differentiation. *Plant Mol.Biol.* **37**, 1069-1078.

van den Berg C. et al. (1995) Cell fate in the Arabidopsis root meristem determined by directional signalling . *Nature* **378**, 62-65.

van den Berg C. et al. (1997) Short-range control of cell differentiation in the Arabidopsis root meristem. *Nature* **390**, 287-289.

van der Zaal E.J. et al. (1991) Promoters of auxin-induced genes from tobacco can lead to auxin- inducible and root tip-specific expression. *Plant Mol.Biol.* **16**, 983-998.

van Haaren M.J. et al. (1987) Overdrive is a T-region transfer enhancer which stimulates T-strand production in Agrobacterium tumefaciens. *Nucleic Acids Res.* **15**, 8983-8997.

Van Sluys M.A., Tempe J., & Fedoroff N. (1987) Studies on the introduction and mobility of the maize Activator element in Arabidopsis thaliana and Daucus carota. *EMBO J.* **6**, 3881-3889.

Vidmar J.J. et al. (2000) Cloning and characterization of a root specific high-affinity sulfate transporter from arabidopsis thaliana(1) [In Process Citation]. *FEBS Lett.* **475**, 65-69.

Wanapu C. & Shinmyo A. (1996) cis-regulatory elements of the peroxidase gene in Arabidopsis thaliana involved in root-specific expression and responsiveness to high-salt stress. *Ann.N.Y.Acad.Sci.* **782**, 107-114.

Wang H., Fowke L.C., & Crosby W.L. (1997) A plant cyclin-dependent kinase inhibitor gene [letter]. *Nature* **386**, 451-452.

Wang K. et al. (1987) Sequence context of the T-DNA border repeat element determines its relative activity during T-DNA transfer to plant cells. *Mol.Gen.Genet.* **210**, 338-346.

Webb K.J. et al. (2000) Mesorhizobium loti increases root-specific expression of a calcium- binding protein homologue identified by promoter tagging in Lotus japonicus [In Process Citation]. *Mol.Plant Microbe Interact.* **13**, 606-616.

Wilson A. et al. (1999) Evidence for the activation of a MAP kinase upon phosphate-induced cell cycle re-entry in tobacco cells. *Physiol.Plant.* **102**, 532-538.

Winicov I. & Bastola D.R. (1999) Transgenic overexpression of the transcription factor alfin1 enhances expression of the endogenous MsPRP2 gene in alfalfa and improves salinity tolerance of the plants. *Plant Physiol* **120**, 473-480.

Winicov I. (2000) Alfin1 transcription factor overexpression enhances plant root growth under normal and saline conditions and improves salt tolerance in alfalfa. *Planta* **210**, 416-422.

Wodak S.J. (1987) Computer-aided design in protein engineering. *Ann.N.Y.Acad.Sci.* **501**, 1-13.

Woo H.H. & Hawes M.C. (1997) Cloning of genes whose expression is correlated with mitosis and localized in dividing cells in root caps of Pisum sativum L. *Plant Mol.Biol.* **35**, 1045-1051.

Woulfe J. et al. (1994) Post-translational processing of the neurotensin/neuromedin N precursor in the central nervous system of the rat--II. Immunohistochemical localization of maturation products. *Neuroscience* **60**, 167-181.

Xie Q. et al. (1996) Plant cells contain a novel member of the retinoblastoma family of growth regulatory proteins. *EMBO J.* **15**, 4900-4908.

Xu Y. et al. (1995) Characterization of a rice gene family encoding root-specific proteins [published erratum appears in Plant Mol Biol 1995 Mar;27(5):1060]. *Plant Mol.Biol.* **27**, 237-248.

Yamamoto E., Karakaya H.C., & Knap H.T. (2000) Molecular characterization of two soybean homologs of Arabidopsis thaliana CLAVATA1 from the wild type and fasciation mutant. *Biochim.Biophys.Acta* **1491**, 333-340.

Yamamoto Y.T. et al. (1991) Characterization of cis-acting sequences regulating root-specific gene expression in tobacco. *Plant Cell* **3**, 371-382.

Zeng Y. et al. (1998) Replication checkpoint requires phosphorylation of the phosphatase Cdc25 by Cds1 or Chk1. *Nature* **395**, 507-510.

Zhang Y.L. et al. (1996) A superactive peptidomimetic analog of a farnesylated dodecapeptide yeast pheromone. *Biochem.Biophys.Res.Commun.* **224**, 327-331.

CLAIMS

1. An isolated nucleic acid comprising a transcriptional regulatory root promoter or a functional fragment of such a promoter and/or comprising a nucleic acid sequence encoding a novel LRR receptor-like kinase (protein) or an immunologically active and/or functional fragment of such a protein (kinase) selected from the group consisting of:

   (a) a nucleic acid comprising the DNA sequence as given in any of SEQ ID NOs 1, 2, 3 or 18, or the complement thereof,

   (b) a nucleic acid comprising the RNA sequences corresponding to any of SEQ ID NOs 1, 2, 3 or 18, or the complement thereof,

   (c) a nucleic acid specifically hybridizing with the nucleotide sequence as defined in (a) or (b),

   (d) a nucleic acid encoding a protein comprising the amino acid sequence as given in SEQ ID NO 4, SEQ ID NO 5 or SEQ ID NO 19,

   (e) a nucleic acid which is degenerated, as a result of the genetic code, to a nucleotide sequence as defined in any of (a) to (d), and which sequence codes for a protein having the amino acid sequence as given in SEQ ID NO 4, SEQ ID NO 19,

   (f) a nucleic acid which is diverging, due to the differences in codon usage between the organisms, to a nucleotide sequence as defined in any of (a) to (d), and which sequence codes for a protein having the amino acid sequence as given in SEQ ID NO 4, SEQ ID NO 19,

   (g) a nucleic acid which is diverging, due to the differences between alleles, to a nucleotide sequence as defined in any of (a) to (d), and which sequence codes for a protein having the amino acid sequence as given in SEQ ID NO 4, SEQ ID NO 19,

   (h) a nucleic acid encoding a fragment of a protein encoded by a DNA sequence as given in SEQ ID NO 1 or SEQ ID NO 3 or encoding a fragment of a protein encoded by a nucleic acid as defined in any one of (a) to (g), wherein said fragment comprises the sequence as represented in SEQ ID NO 5, and,

(k) a nucleic acid encoding a protein as defined in SEQ ID NO 4, SEQ ID NO 19 or a nucleic acid as defined in any one of (a) to (h), said nucleic acid interrupted by intervening DNA,

provided that said nucleic acid is not one of the nucleic acids as deposited under the GenBank Accession numbers AB017061, AV538507, AV520681 or AQ966419.

2. An isolated nucleic acid comprising a transcriptional regulatory root promoter comprising the sequence as given in any of SEQ ID NOs 1, 2 or 18.

3. An isolated nucleic acid comprising a transcriptional regulatory root promoter consisting of the sequence as given in SEQ ID NO 2 or 18, or a functional part of said sequence which is able to regulate gene expression in a root specific manner, provided that said nucleic acid is not one of the nucleic acids as deposited under the GenBank Accession numbers AV538507 or AV520681.

4. An isolated nucleic acid comprising a transcriptional regulatory root specific promoter comprising:

(a) at least a functional part of the DNA sequence as given in any of SEQ ID NO 1, 2 or 18, and,

(b) a second transcriptional regulatory sequence, optionally a regulatory promoter sequence not normally exhibiting root-specificity.

5. An isolated nucleic acid according to claim 4 wherein said second transcriptional regulatory promoter sequence is the ubiquitin promoter.

6. An isolated nucleic acid selected from the group consisting of:

(a) a nucleic acid consisting of at least part of the DNA sequence as given in SEQ ID NO 1 or 3, or the complement thereof,

(b) a nucleic acid encoding a protein as given in SEQ ID NO 4 or 19 or encoding a fragment of said protein, wherein said fragment comprises the sequence as represented in SEQ ID NO 5, and

(c) a nucleic acid encoding a protein with an amino acid sequence which is at least 65 % identical to the protein as given in SEQ ID NO 4 or 9, wherein said amino acid sequence comprises the sequence as represented in SEQ ID NO 5,

or a homologue or a derivative of said protein, or an immunologically active and/or functional fragment thereof.

16. A polypeptide as defined in claim 15 being a root CLAVATA homologue or a functional homologue thereof

17. An isolated polypeptide comprising the sequence represented in SEQ ID NO 5 encodable by a nucleic acid of any of claims 1 to 7, or a homologue or a derivative thereof, or an immunologically active and/or functional fragment thereof.

18. A method for producing a polypeptide according to any of claims 15 to 17 comprising culturing a host cell of claim 13 or 14 under conditions allowing the expression of the polypeptide and recovering the produced polypeptide from the culture.

19. An antibody specifically recognizing an LRR receptor-like kinase (protein) of any of claims 15 to 17 or recognizing immunologically active parts or specific epitopes thereof.

20. A method for the production of transgenic plants, plant cells or plant tissues comprising the introduction of a nucleic acid molecule according to any of claims 1 to 7 in an expressible format or a vector according to any of claims 10 to 12 in said plant, plant cell or plant tissue.

21. The method of claim 20 further comprising regenerating a plant from said plant cell.

22. A transgenic plant cell comprising a nucleic acid sequence of any of claims 1 to 7 which is operably linked to regulatory elements allowing transcription and/or expression of said nucleic acid in plant cells or obtainable by a method of claim 20 or 21.

23. The transgenic plant cell of claim 22 wherein said nucleic acid of any of claims 1 to 7 is stably integrated into the genome of said plant cell.

24. A transgenic plant or plant tissue comprising plant cells of claim 22 or 23.

25. A harvestable part of a plant of claim 24.

26. The harvestable part of claim 25 which is selected from the group consisting of seeds, leaves, fruits, stem cultures, rhizomes, tubers and bulbs.

27. The progeny derived from any of the plants or plant parts of any of claims 24 to 26.

characterised in that said nucleic acid encodes a novel LRR receptor-like kinase protein or an immunologically active and/or functional fragment of such a protein, and further provided that said nucleic acid is not one of the nucleic acids as deposited under the GenBank Accession numbers AB017061 or AQ966419.

7. An isolated nucleic acid according to any of claims 1 to 6 which is DNA, cDNA, genomic DNA, synthetic DNA, or RNA wherein T is replaced by U.

8. A nucleic acid molecule of at least 15 nucleotides in length hybridizing specifically with a nucleic acid of any of claims 2 to 7.

9. A nucleic acid molecule of at least 15 nucleotides in length specifically amplifying a nucleic acid of any of claims 2 to 7.

10. A vector comprising a nucleic acid sequence according to any of claims 1 to 7.

11. A vector according to claim 10 which is an expression vector wherein said nucleic acid sequence encoding a novel LRR receptor-like kinase or an immunologically active and/or functional fragment thereof, is operably linked to one or more control sequences allowing the expression in prokaryotic and/or eukaryotic host cells.

12. A vector according to claim 10 wherein said vector comprises at least part of a nucleic acid according to any of claims 2 to 7 and wherein said transcriptional regulatory root promoter or functional parts thereof is operably linked to one or more genes of interest.

13. A host cell containing a nucleic acid molecule according to any of claim 1 to 7 or a vector according to any of claims 10 to 12.

14. The host cell according to claim 13, wherein said host cell is a bacterial, insect, fungal, plant or animal cell.

15. An isolated LRR receptor-like kinase comprising one of the polypeptides selected from the group consisting of:

(a) a polypeptide as given in SEQ ID NO 4 or SEQ ID NO 19,

(b) a polypeptide comprising the amino acid sequence as given in SEQ ID NO 5, and,

(c) a polypeptide encoded by a nucleic acid as given in SEQ ID NO 1 or 3,

28. A method for conferring root-specificity, root-meristem-specificity, root-vascular-tissue-specificity, root-endodermis-specificity, root-cortex-specificity, root-epidermis-specificity, root-quiescent-center-specificity, and/or abundant expression in root and/or early expression in root to other promoter sequences comprising the fusion at least a functional part of the DNA sequence as given in SEQ ID NO 1, 2 or 18 to a second transcriptional regulatory promoter sequence normally not exhibiting root-specificity.

29. A method for root-specific expression of a gene(s) of interest comprising operably linking of said gene(s) of interest to a transcriptional regulatory root-specific promoter as defined in any of claims 2 to 5.

30. A method for modifying cell fate and/or plant development and/or plant morphology and/or plant biochemistry and/or plant physiology comprising the modification of expression in the meristem of the main and the lateral roots, in the vascular tissue of the root or in the lateral root primordia of a gene(s) of interest operably linked to a transcriptional regulatory root-specific promoter as defined in any of claims 2 to 5.

31. A method for phytoremediation comprising the expression of a gene(s) of interest under the control of a transcriptional regulatory root-specific promoter as defined in any of claims 2 to 5.

32. A method for environmental remediation comprising the expression of a gene(s) of interest under the control of a transcriptional regulatory root-specific promoter as defined in any of claims 2 to 5.

33. A method for enhancing plant growth and crop yield comprising the expression of a gene(s) of interest under the control of a a transcriptional regulatory root-specific promoter as defined in any of claims 2 to 5.

34. A method for conferring enhanced resistance to pathogens which attack the belowground plant tissue comprising the expression of a gene(s) of interest under the control of a transcriptional regulatory root-specific promoter as defined in any of claims 2 to 5.

35. A method for enhancing freezing tolerance in plants comprising the expression of a gene(s) of interest under the control of a transcriptional regulatory root-specific promoter as defined in any of claims 2 to 5.

36. A method for enhancing resistance of a plant to drought and/or high salt comprising the expression of a gene(s) of interest under the control of a transcriptional regulatory root-specific promoter as defined in any of claims 2 to 5.

37. A method for the production of light-sensitive proteins comprising the expression of a gene encoding said light-sensitive protein under the control of a transcriptional regulatory root-specific promoter as defined in any of claims 2 to 5.

38. A method for root specific gene silencing comprising the expression of an antisense nucleotide sequence to the gene of interest, a gene silencing construct or a ribozyme under the control of a transcriptional regulatory root-specific promoter as defined in any of claims 2 to 5.

39. A method for stimulating root meristem formation comprising the expression of a gene that influences the proliferation and/or differentiation of the root meristem under the control of a transcriptional regulatory root-specific promoter as defined in any of claims 2 to 5.

40. A method for stimulating root meristem formation comprising the expression in particular cells, tissues or organs of a plant, of a nucleic acid encoding an LRR receptor-like kinase operably linked to a plant-operable promoter sequence.

41. A method stimulating root meristem formation comprising the expression in particular cells, tissues or organs of a plant, of a nucleic acid encoding an LRR receptor-like kinase as defined in any of claims 15 to 17 operably linked to a plant-operable promoter sequence.

42. A method stimulating root meristem formation comprising the expression in particular cells, tissues or organs of a plant, of a nucleic acid encoding an LRR receptor-like kinase under the control of a transcriptional regulatory root promoter as defined in any of claims 2 to 5.

43. A method for stimulating root meristem formation comprising the expression in particular cells, tissues or organs of a plant, of a nucleic acid encoding a polypeptide as defined in any of claims 15 to 17 under the control of a transcriptional regulatory root promoter as defined in any of claims 2 to 5.

44 A method for root-meristem maintenance comprising the expression of a gene that influences the proliferation and/or differentiation of the root meristem under the control of a transcriptional regulatory root-specific promoter as defined in any of claims 2 to 5.

45 A method for root-meristem maintenance comprising the expression in particular cells, tissues or organs of a plant, of a nucleic acid encoding an LRR receptor-like kinase operably linked to a plant-operable promoter sequence.

46 A method for root-meristem maintenance comprising the expression in particular cells, tissues or organs of a plant, of a nucleic acid encoding an LRR receptor-like kinase as defined in any of claims 15 to 17 operably linked to a plant-operable promoter.

47 A method for root-meristem maintenance comprising the expression in particular cells, tissues or organs of a plant, of a nucleic acid encoding an LRR receptor-like kinase under the control of transcriptional regulatory root promoter as defined in any of claims 2 to 5.

48 A method for root-meristem maintenance comprising the expression in particular cells, tissues or organs of a plant, of a nucleic acid encoding a polypeptide as defined in any of claims 15 to 17 under the control of a transcriptional regulatory root promoter as defined in any of claims 2 to 5.

49 A method for enhancing root formation and/or root growth comprising the expression in particular cells, tissues or organs of a plant, of a nucleic acid encoding an LRR receptor-like kinase operably linked to a plant-operable promoter sequence.

50 A method for enhancing root formation and/or root growth comprising the expression in particular cells, tissues or organs of a plant, of a nucleic acid encoding an LRR receptor-like kinase as defined in any of claims 15 to 17 under the control of a plant operable promoter sequence.

51 A method for enhancing root formation and/or root growth comprising the expression in particular cells, tissues or organs of a plant, of a nucleic acid encoding an LRR receptor-like kinase under the control of a transcriptional regulatory root promoter as defined in any of claims 2 to 5.

52. A method for enhancing root-formation and/or root growth comprising the expression in particular cells, tissues or organs of a plant, of a nucleic acid encoding a polypeptide as defined in any of claims 15 to 17 under the control of a transcriptional regulatory root promoter as defined in any of claims 2 to 5.

53. A method for enhancing overall growth and yield comprising the expression in particular cells, tissues or organs of a plant, of a nucleic acid encoding an LRR receptor-like kinase operably linked to a plant-operable promoter sequence.

54. A method for enhancing overall growth and yield comprising the expression in particular cells, tissues or organs of a plant, of a nucleic acid encoding an LRR receptor-like kinase as defined in any of claims 15 to 17 operably linked to a plant operable promoter sequence.

55. A method for enhancing overall growth and yield comprising the expression in particular cells, tissues or organs of a plant, of a nucleic acid encoding an LRR receptor-like kinase under the control of a transcriptional regulatory root promoter as defined in any of claims 2 to 5.

56. A method for enhancing overall growth and yield comprising the expression in particular cells, tissues or organs of a plant, of a nucleic acid encoding a polypeptide as defined in any of claims 15 to 17 under the control of a transcriptional regulatory root promoter as defined in any of claims 2 to 5.

57. A method for modifying cell fate and/or plant development and/or plant morphology and/or plant biochemistry and/or plant physiology comprising the modification of expression in particular cells, tissues or organs of a plant, of a nucleic acid encoding an LRR receptor-like kinase operably linked to a plant-operable promoter sequence.

58. A method for modifying cell fate and/or plant development and/or plant morphology and/or plant biochemistry and/or plant physiology comprising the modification of expression in particular cells, tissues or organs of a plant, of a nucleic acid encoding an LRR receptor-like kinase as defined in any of claims 15 to 17 operably linked to a plant-operable promoter sequence.

59. A method for modifying cell fate and/or plant development and/or plant morphology and/or plant biochemistry and/or plant physiology comprising the modification of expression in particular cells, tissues or organs of a plant, of a nucleic acid encoding

an LRR receptor-like kinase under the control of a transcriptional regulatory root promoter as defined in any of claims 2 to 5.

60. A method for modifying cell fate and/or plant development and/or plant morphology and/or plant biochemistry and/or plant physiology comprising the modification of expression in particular cells, tissues or organs of a plant, of a nucleic acid encoding a polypeptide as defined in any of claims 15 to 17 under the control of a transcriptional regulatory root promoter as defined in any of claims 2 to 5.

61. A method to confer pathogen resistance to a plant comprising the expression in particular cells, tissues or organs of a plant, of a nucleic acid encoding an LRR receptor-like kinase operably linked to a plant-operable promoter sequence.

62. A method to confer pathogen resistance to a plant comprising the expression in particular cells, tissues or organs of a plant, of a nucleic acid encoding an LRR receptor-like kinase as defined in any of claims 15 to 17 operably linked to a plant-operable promoter sequence.

63. A method to confer pathogen resistance to a plant comprising the expression in particular cells, tissues or organs of a plant, of a nucleic acid encoding an LRR receptor-like kinase under the control of a transcriptional regulatory root promoter as defined in any of claims 2 to 5.

64. A method to confer pathogen resistance to a plant comprising the expression in particular cells, tissues or organs of a plant, of a nucleic acid encoding a polypeptide as defined in any of claims 15 to 17 under the control of a transcriptional regulatory root promoter as defined in any of claims 2 to 5.

65. A method for identifying and obtaining proteins interacting with a (LRR receptor-like kinase) polypeptide of any of claims 15 to 17 comprising a screening assay wherein a polypeptide of any of claims 15 to 17 is used or expressed.

66. A method according to claim 65 comprising a two-hybrid screening system wherein a nucleic acid encoding a polypeptide of any of claims 15 to 17 as a bait and a cDNA library as prey are expressed.

67. A method for modulating the interaction between an LRR receptor-like kinase (protein) as defined in any of claims 15 to 17 and interacting proteins obtainable by a method according to claim 65 or 66.

68. A method for identifying and obtaining compounds interacting with an LRR receptor-like kinase (protein) comprising the steps of:

(a) providing a two-hybrid screening system wherein a polypeptide of any of claims 15 to 17 and a protein interacting with said LRR receptor-like kinase (protein) or an interacting protein obtainable by a method of claim 65 or 66 are expressed,

(b) interacting said compound with the complex formed by the expressed proteins as defined in a),

(c) detecting a second complex, wherein the presence of said second complex identifies a compound which specifically binds to one of said polypeptides or said second complex, and

(d) identifying the compound.

69. A method for identifying compounds or mixtures of compounds which specifically bind to a polypeptide of any of claims 15 to 17, comprising:

(a) combining a polypeptide of any of claims 15 to 17 with said compound or mixtures of compounds under conditions suitable to allow complex formation, and,

(b) detecting complex formation, wherein the presence of a complex identifies a molecule which specifically binds said polypeptide.

70. Use of a molecule identifiable by means of a method of claim 68 or 69 as a plant growth regulator or herbicide.

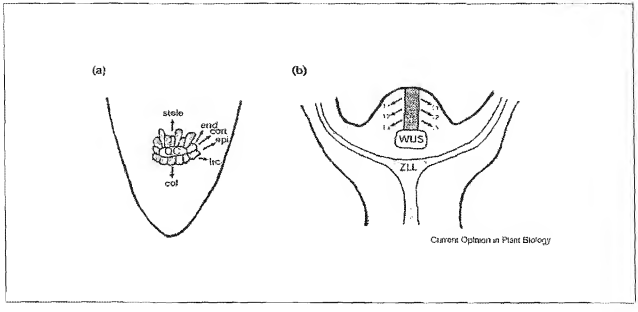
71. A method for the production of a plant growth regulator or herbicide composition comprising the steps of the method of claim 68 or 69 and formulating the compounds obtained from said steps in a suitable form for the application in agriculture or plant cell or tissue culture

72. Use of a nucleic acid molecule of any of claims 1 to 7, a vector of claim 10 to 12, a polypeptide of claim 15 to 17 or an antibody of claim 19 for modifying cell fate and/or

plant development and/or plant morphology and/or plant biochemistry and/or plant physiology.

73. Diagnostic composition comprising at least one nucleic acid molecule of any of claims 1 to 7, vector of claim 10 to 12, polypeptide of claim 15 to 17 or antibody of claim 19.

Figure 1


(a)
stele
endd
cort
epi
QC
lrc
col
(b)
WUS
ZLL
Current Opinion in Plant Biology

Primers used in experiments described

SEQ ID No.	Primer	Sequence (5'-3')	Details
SEQ ID No. 6	LRR1	AAYAAYTTMASYGGTKMDWTHCC	Degenerate LRR primer
SEQ ID No. 7	LRR2	CARYTHAMCGGWGAAATHCC	Degenerate LRR primer
SEQ ID No. 8	LRR3	MAYYTSAIYGGAHCTATHCC	Degenerate LRR primer
SEQ ID No. 9	KIN1	THCCWTTWGSCATGTANTCATA	Degenerate kinase primer
SEQ ID No. 10	KIN2	AYTCDGGDGCHAHGTADCC	Degenerate kinase primer
SEQ ID No. 11	KIN3	MAYHCCRAARCTRTAVACATC	Degenerate kinase primer
SEQ ID No. 12	RCH1F	CGATCAGACACAAGAACAT	*RCH1* specific primer
SEQ ID No. 13	RCH1R	AGCAATGGTGTTGGAAGAA	*RCH1* specific primer
SEQ ID No. 14	UBIF	TGCAGATCTTYGTGAAGAC	Compl. to 5' end of repeating ubiquitin unit
SEQ ID No. 15	UBIR	GACTCCTTCTGGATGTTG	Compl. to 3' end of repeating ubiquitin unit
SEQ ID No. 16	M13F	TGTAAAACGACGGCCAGT	Compl. to pBS sequence
SEQ ID No. 17	PRCH1R	CACACAGGATCCAATCGGCATTT GCAAAGACATA	Compl. to *RCH1* promoter, includes small coding sequence and BamHI site

N=A+C+G+T; Y=C+T; M=A+C; S=C+G; K=T+G; W=A+T; R= A+G; D=A+T+G; H=A+T+C; V=A+C+G; I=deoxyinosine

Figure 3

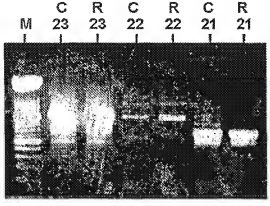


Figure 4

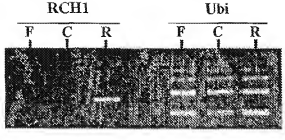


Figure 5-1 SEQ ID NO 1 : 7438 bp

CTCGAGTTTCAGATGTTTCTATTAAATAAGTAATGAAAGCAGCTCTAATTCTCCCAAAATTCACAAGTATAGAAAGATAG
AAAGATAGAAACAGAGTTATTGCATGGAAGCTGAAAATCATACAGCATGGTGGATATAACCAATGAGACACAAAACAATC
ACGGCATGTGTTATAGAACAACACTTATCATCTGTATTAACATAAAGCAAAAGTTCCTATACGACATAGTCCAAAGTTAC
AGACCATAAGCACACCTTATGATTTGTTGTACCCTCTAAAACACAGGATGATGACAAGAAGCAAATTTGCCCAAATTTAC
TGAGGGAAAAACGAGGGCTTAACGCTCTCATATCTCAAACAAAAACTTCTCAAACAGTTTCATGTGTTCAGATTGGAGGG
CAATGGCAACGGATAAGCTGCCATCATTAGTAGGACTTGGTAGCACAAAAGACAAACCCTCGTATGGAATTCCACCAGGT
CCCATAAAGATAGGACGACCCCAACCAAAGTCTGCATCATAAATAGGTAATCTAACCCAGCTTGTGATTCCCAAATTTGG
GCACTTGTAGGTATGTGCACCGCGGACAAGGGCTGACAGATCAGGCTGCATCTCCAGGTAGTCAAGAGCTGACCTCAGAT
AGTTATCATCCATACGAACCAAGAAATCATGAATCTGTCCTGCAGCATACCATGTTGGCTTAGATAACAAATCCCCTGCA
ACAGCCAATGGTGTTGCAGTGAATATCACATTCCCAAAGTAACCAGGAGGCAGCTGCGGACGTAGTCTAGACCTTCCATC
AGTTGCAATGTACAGTTTCGTCTCTTGGTCGTTTGGAAGCCCTCGCGCCTTTCCCACTGATCTCCACACATGCCCTGCCA
ACATCTCGTATGAGCTGTAGCTGACAGTGTTCCCATCCTCCTTGGATTTCGCCTTAAGAGCAACAAGCTGGTCTCGTGTT
AATTTGAATATAGAGACAGTGGTATTCTCAGGTCCTGATTTAGACGGATCAAGAGGTATCTTCATACTTGGTGCAGGCTG
ATATTCAACATGATGAAAAGCAGGCTGTGGCGGGTCCCTAGCTCGGAGGAGTGTTCGATCAATGAAAGGTGGAATGGTTA
GGTCAAGACCACGAGCCATATCAGACCATGTGTTGATAAAATGAAGACCAGAGAAACCATCTGCCGCGTGATGTTGCATC
CCAACCCCAAGTGAAGCTCCCCCACATTTAAAGAAAGTCACCTATATCATCAAACAATATCATTACATAACTAAAAACAG
AACATACATCATCAAAGAATATCATTACAGAACTAAAAAACAGAACAACACAAGCATTCAATAAATTAAAAAGGGAAAG
ACTTTAAATTAAGCTGACCTGCAAAACGAGAAGCGGGAAAGAGTGAATGCCAGCGGAGTGATCAACTTCGGGAATAAGCT
GACGGAGATTAAGGGTAGGAGCAAAATCACCAAAATCATCGATAACAGAAGGAGTATCAGCCACAACGAAGAGAACACCA
GCACCGTTACAATCGATCTCAATACGACCATCATCGTCTCTCTTCAAGCGACCAGCCATAGGGTAAAAAGGGACAAGGGC
TTTGGAAAGAGCTTCCTTCATGACCTGAGGGTCAAAGAAATTGGAAGCGCCGGTGGGTCTGTAGAAGTAGACACTAGGGG
TATGGAATCTGGGGATGACAAGGTCGACGTTGGAGTTCCAAAGATTAGTGATTGGTGTCTCGGTGGCAGGCCGGACCATG
GTGGAATCTCTGATGTTAATTTTCATATTTTTAAGAAAGATCAAACGAACAAACAAATTTTGTTTTGAGTTTGATCTGAG
AAATTGTGTGAGAAGCGAGGAGGGGTAGATATAAGTGGGTGTGAACAAAAAAGGAAAGGTCGTAATCCTTCACCTACCGC
CCCACGACCCCGCCTACTCCATCTGTCGACCAACTCTCTCAGAAGGTTTATCATTGGTCCATCTCATCACTCGGTCCTAC
TTTTCTATCCGGGGTTGGTAAACCCGTTCCCTCTGTCCTAAGGCCCAAATAATACTAATGGGCTTAAATAGTATAAAGTT
CGAATGAAGAAGAAGGAAAAAAAATTGACAAGTTGATTGTTATAGCTACCGTACAAGTTAGTCAATCGAATGCCAAAGTT
TGGCTTAACTTAACTCATTACCACGTAGCAGAGCATAACAACAACCACCAATAACATCTAACGGCGTAAACTCACGTGTT
AAGAGAATTACGTCAGAATCCAAGAAAGTCTTTTTTTCAGTACACCACAAATAAATTATGTTGAATTTATATCCGAACTA
TACTATATTATATGAAAACATACCTCAATTAGTCATAAATTTCACATTTCATTTTAATTTATTAACTAATGGTTGTTGTT
GCGGATAGCTTTTCTTTGCTTTCTGTTCTAAAATTATTTCTTAGCCATTATTTGATGATCCAACTTGGTTTGAGGATCTA
ATAAAAATTCAATTTCTTTGAAGCATTCATTGGCTTAATGATCCAAACCTTTTGTCCTCCCACTCTCTAAGAAAATAAAG
CAAAAGCAAAACTAAAAGAAAAGCAACTAGTACCGTGGGTTGGTCAATGGTAACCCATGCTACCAAATACCAATCAAAGC
TTAAATGACAATTTAGTACCTTGGGTTGGTCATGATTTAGAGCGGAACAAATATACCATACATCAAACGAGGATATACAG
AGAAAATTCATGGAAGTATGGAATTTAGAGGACAATTTCTCTTCTGGGCTACAACGGACCGGCCCATTCGCTCATTTACC
CAGAGGTATCGAGTTTGTGGACTTTTGATGCCGCTAGAGACTATTGGCATCGGATTGAAAAAAATGTTTACTTCGTTGTT
AACAATTTTCTGAATGCAATATTTTCCTTGTCATGAATATTTAAACTTGTTATTACTTTCTTTTAGCTTAGGTGTGGACA
ATTATGGAGTTTACTTCAAACGAGGAAGAATCTTAAACGCTCGGTTCAGGTCTCGAAAACAAACCAACTCACAATCCTGA
CTTAATTGAGGAAAACAATGCAAAACCACATGCATGCTTCCATATTTCTATCATAATCTTATAAGAAAAAACACTACTAA
GTGAAATGATTCTGTATATATATATAACCAATGCCTTTTGTTTTGTGATATTTTATGTATATATAACTATTGACTTTTGTCA
TCTATGGATAGTGTCTCGGGCTCTTGGCAAACATATTTCAAAGAAAAGTTAATGACTGTAATTAATTAATCTGAAGCTAG
AAACAGAACCCCGAGGTAAAAGAAAAAGACAGAGCACATGAAGTTTAGTACTTTTATATATTTAATATATCATTCTTTCT
TATTGCTTATCTCTAAAGCAAAAACTTCCCTAAACCCTAAGCCAAAGGACTCAGATCGATGCAGAACCAAGAAGGCTTGT
TTTGGATTTGAGAGCCAAATGCAAAGAAAAAAACTCTTATGTCTTTGCAAATGCCGATTCCAAGAAAAAAAGCCTTAACT
GTTTCTCATTTTTCCATTACATTGTCTCTGTTTTTAGCTTTCTTCATCTCCTCGACCTCTGCATCAACCAATGAAGTCTC
AGCTTTGATTTCTTGGCTTCACAGCTCGAACTCGCCGCCACCGTCAGTCTTCTCCGGCTGGAATCCTTCTGATTCTGACC
CATGTCAATGGCCTTACATTACTTGTTCCTCCTCAGACAACAAACTCGTTACAGAGATCAATGTCGTCTCTGTTCAGTTA
GCTTTACCTTTCCCTCCTAACATTTCTTCGTTTACTTCACTTCAGAAACTCGTTATCTCCAACACTAATCTCACCGGAGC
TATATCTTCTGAGATCGGAGATTGTTCGGAGCTTATAGTTATCGATTTAAGCTCAAATAGTCTTGTTGGTGAAATACCTT
CGAGTCTAGGGAAGCTCAAGAATCTTCAAGAACTTTGTTTAAACTCCAATGGCCTCACAGGAAAGATCCCACCGGAACTC
GGCGACTGCGTTAGCCTCAAGAATCTTGAGATTTTCGATAACTACTTATCCGAGAATCTCCCGTTGGAGCTCGGAAAGAT
CTCGACTCTTGAGAGCATAAGAGCCGGAGGAAACTCAGAGCTTTCAGGGAAGATCCCGGAGGAGATCGGAAACTGTAGGA
ATCTCAAGGTCTTAGGGTTAGCAGCAACGAAAATATCCGGTTCTTTACCTGTTTCGTTGGGTCAACTAAGCAAGCTCCAG
TCTTTATCTGTCTATTCTACAATGCTCTCAGGTGAGATCCCTAAAGAGCTTGGAAACTGCTCTGAACTTATCAATCTGTT
TCTATACGACAATGACTTGTCCGGCACGCTTCCAAAAGAACTAGGAAAGCTTCAAAACCTAGAGAAGATGCTTCTATGGC
AGAATAATCTCCACGGACCTATCCCTGAGGAGATTGGATTCATGAAAAGCTTAAACGCCATTGATCTCTCTATGAACTAT
TTCTCAGGAACCATCCCTAAATCGTTCGGTAACTTGTCGAATCTCCAAGAGCTTATGCTTAGCAGCAACAACATCACCGG
GTCAATACCTTCGATTCTAAGTAACTGCACAAAGCTTGTTCAGTTCCAGATTGACGCCAATCAGATTTCAGGTTTGATTC
CACCAGAGATTGGATTGCTTAAGGAGCTCAACATCTTCTTGGGATGGCAGAATAAGCTAGAAGGGAATATTCCGGACGAG
TTAGCTGGTTGTCAGAATCTTCAAGCTCTTGACTTGTCACAGAATTATCTCACCGGATCATTACCTGCAGGTTTGTTTCA
GCTTCGTAATCTGACTAAGCTCTTGCTTATATCTAATGCCATCTCCGGTGTTATCCCACTGGAGATCGGGAACTGCACTT
CGCTTGTCAGATTAAGGCTTGTCAATAACAGAATCACCGGAGAAATCCCTAAAGGTATTGGATTTCTTCAGAACCTTAGC
TTCCTCGACTTGTCTGAGAACAATCTCTCCGGTCCAGTTCCTTTGGAGATAAGTAATTGTAGACAACTCCAGATGCTTAA
CTTGAGCAACAACACTCTTCAAGGTTATCTTCCTCTGTCTTTATCATCGTTAACAAAGCTTCAAGTACTTGATGTCTCTT
CCAATGACTTGACTGGCAAGATACCTGATAGTCTAGGTCATCTCATTTCACTCAACCGGCTCATTCTAAGTAAGAATTCT
TTCAATGGAGAAATCCCTTCGTCTCTCGGTCACTGCACGAATCTTCAGCTTCTTGATCTCAGCAGTAATAACATCTCTGG
AACTATACCGGAGGAGCTTTTTGACATCCAAGATCTAGACATTGCGTTAAACTTGAGCTGGAATTCATTAGATGGCTTTA
TCCCGGAAAGGATTTCCGCGCTTAACCGATTATCCGTCCTCGATATTTCGCATAACATGCTTTCAGGCGACCTTTCTGCG
CTGTCCGGTCTAGAAAACTTGGTTTCTCTGAACATCTCTCATAACAGATTCTCAGGTTATCTTCCAGATAGTAAAGTGTT
TAGACAGCTGATAGGAGCAGAGATGGAAGGAAACAATGGACTCTGTTCCAAAGGTTTCAGGTCTTGCTTTGTAAGTAACA
GTTCACAGTTAACTACACAGCGTGGTGTGCACTCACATAGACTCAGGATAGCCATTGGATTGCTAATCAGCGTGACAGCG

GTTCTAGCGGTATTAGGCGTGTTGGCGGTTATACGAGCTAAACAAATGATTCGAGATGATAACGATTCAGAGACTGGAGA
AAATCTATGGACATGGCAATTCACACCTTTTCAGAAACTCAACTTCACAGTCGAACATGTACTCAAGTGTTGGTAGAAG
GTAATGTTATAGGGAAAGGTTGCTCCGGGATAGTGTACAAAGCTGAAATGCCTAACAGAGAAGTCATCGCGGTGAAAAAG

Figure 5-2

CTCTGGCCAGTGACAGTACCTAATCTGAATGAGAAAACTAAGTCATCAGGAGTTCGAGACTCATTCTCAGCTGAAGTAAA
AACACTTGGATCGATCAGACACAAGAACATTGTACGCTTCTTGGGATGTTGTTGGAACAAGAACACTAGACTTCTTATGT
ATGATTATATGTCAAATGGGAGTTTGGGAAGTTTGCTTCACGAAAGGAGCGGTGTATGTAGCTTAGGATGGGAAGTGAGG
TACAAGATTATACTTGGTGCAGCTCAGGGTTGGCTTACTTGCACCATGACTGTGTTCCTCCATTGTTCATAGAGACAT
CAAGGCAAATAATATTCTGATTGGCCCTGATTTTGAACCTTACATTGGAGATTTCGGACTCGCTAAGCTTGTTGATGATG
GCGACTTTGCTCGTTCTTCCAACACCATTGCTGGTTCCTATGGTTACATAGCTCCAGGTACTCCCCAAATTTCTGTTTTG
CTGCTAATTAGTTAACTTGATACATCATAAGTTACTAGTAGTCTTAGCCTAGTAATGGTAGAAATTCAGGTTTATAGTTT
GGTTTTGGTTTTTGGCATTTTGAAAAGAATAGTATTTTGGTTCAGTTGATTTTCGATTTGGTTTAAAATTTTAGAAAAAA
AATGGTAGAAGAAATAATTTTTGATTAATATTATATCATCTACTGTATACGATTTAAAAAAAAAATTGATTTGGTTTGGT
CTCGATATTTTGCTTGTAGCTTGGTTCAATATTTTACTATATATATGATTCATGACTTTCATACTAATCAATATTTGTT
CGGTTTCGTTTTGATTAGTCTGAAAGTGAATCATATAGTTAAAATATTGACCAATCTATCATATACTTACAGCTACAAG
ATTTTCGATGTAATTTTCGTCTAAGCATTTTGTTTTATATGTTCATCCTTAGTAACTAACAACAAATATCAATTGGGA
AGCATAACGTTATTATATTCCCATAAGGCTATAACACTATAATTATGCAGAATACGGATACTCAATGAAGATAACAGAGAA
AAGCGACGTGTACAGCTATGGAGTCGTAGTGCTCGAGGTACTAACGGGTAAGCAACCAATCGATCCAACGATACCAGATG
GACTCCACATAGTGGACTGGGTCAAGAAAATCAGAGACATACAAGTAATCGACCAAGGACTACAAGCAAGACCAGAGTCA
GAGGTTGAAGAGATGATGCAAACGCTAGGAGTCGCGCTTTTATGCATTAATCCAATACCCGAAGACAGGCCTACAATGAA
AGATGTGGCTGCTATGCTTAGTGAGATATGTCAAGAGAGAGAGGAATCGATGAAAGTTGATGGTTGCTCGGGAAGTTGTA
ACAATGGAAGAGAACGTGGCAAAGATGATTCGACTTCATCGGTTATGCAACAAACGGCTAAGTATTTGAGAAGTAGTAGC
ACGAGTTTCTCTGCGTCTTCTTTGCTTTACTCTTCTTCTTCTTGCTACTTCTAATGTTAGACCAAATCTTAAATAA

Figure 6 SEQ ID NO 2

```
1    CTCGAGTTTCAGATCTTTCTPATTAAATAAGTAATGAAAGC AGCTCTAATTCTCCCAAAAPCACAAGTATAGAAAGATAG
81   AAAGATAGAAACAGAGTTATTGCATGGAAGCTGAAAATCA TACAGCATGGTGGATATAACCAATGAGACACAAAACAATC
161  ACGGCATGTGTTATAGAACAACACTTATCATCTGTATTAA CATAAAGCAAAAGTTCCTATACGACATAGTCCAAAGTTAC
241  AGACCATAAGCACACCTTATGATTTGTTGTACCCTCTAAA ACACAGGATGATGACAAGAAGCAAATTTGCCCAAATTTAC
321  TGAGGGAAAAACGAGGGCTTAACGCTCTCATATCTCAAAC AAAAACTTCTCAAACAGTTTCATGTGTTCAGATTGGAGGG
401  CAATGGCAACGGATAAGCTGCCATCATTAGTAGGACTTGG TAGCACAAAAGACAAACCCTCGTATGGAATTCCACCAGGT
481  CCCATAAAGATAGGACGACCCACAACAAGCTTCGCATCAT AAATAGGTAATCTAACCCAGCTTGTGATTCCAAATTTGG
561  GCACTTGTAGGTATTGTCACCGCGGACTAGGGCTGACAGA TCAGGCTGCATCTCCAGGTAGTCAAGAGCTGACCTCAGAT
641  AGTTATCATCCATACGAACCAAGAAATCATGAATCTGTCC TGCAGCATACCATGTTGGCTTAGATAACAAATCCCCTGCA
721  ACAGCCAATGGTGTTGCAGTGAATATCACATTCCCAAAGT AACCAGGAGGCAGCTGCGGACGTAGTCTAGACCTTCCATC
801  AGTTGCAATGTTGTACAGTTTCGTCTCTTGGTCGTTTGGAAGC CCTCGCGCCTTTCCCACTGATCTCCACACATGCCCTGCCA
881  ACATCTCGTATGAGCTGTAGCTGACAGTGTPCCCATCCTC CTTGGATTTCGCCTTAAGAGCAACAGCTGGTCTCGTGTT
961  AATTTGAATAATAGAGACAGTGGTATTCTCAGGTCCTGATT TAGACGGATCAAGAGGTATCTTCATACTTGGTGCAGGCTG
1041 ATATTCAACATGATGAAAAGCAGGCTGTGGCGGGTCCCTA GGTCGGAGGAGTGTTCGATCAATGAAAGGTGGAATTGTTA
1121 GGTCAAGACACGAGCCATATCAGACCATGTGTTGATAAA ATGAAGACCACAGAGAAACCACTCGCCGGTGATGTTGCATC
1201 CCAACCCAAGTGAAGCTCCCCCACATTTAAAGAAAGTCA CCTATATCATCAAACAATATCATTACATAACTAAAACAG
1281 AACATACATCATCAAAGAATATCATTACAGAACTAAAAAA CAGAACAACACAAGCATTCAATAAATTAAAAAGGGAAAG
1361 ACTTTAAATTAAGTCAGACCTGCAAAACGAGAAGCGGGAAA GAGTGAATGCCAGCGGAGTGATCAACTTCGGGAATAAGCT
1441 GACGGAGATTAAGGGTAGGAGCAAAATCACCAAAATCATC GATAACAGAAGGAGTATCAGCCACAACGAAGAGAACACCA
1521 GCACCGTTACAATCGATCTCAATACGACCATCATCGTCTC TCTTCAAGCGACCAGCCATAGGGTAAAAGGGACAAGGGC
1601 TTTGGAAAGAGCTTCCTTCATGACCTGAGGGTCAAAGAAA TTGGAAGCGCCGGTGGGTCTGTAGAAGTAGACACTAGGGG
1681 TATGGAATCTCGGGATGACAAGGTCGACGTTGGAGTTCCA AAGATTAGTGATTGGTGTCTCGGTGGCAGGCCGGACCAATG
1761 GTGGAATCTCTGATGTTAATTTTCATATTTTTAAGAAAGA TCAAACGAACAAACAAATTTGTTTTGAGTTTGATCTGAG
1841 AAATTGTGTGAGAAGCGAGGAGGGGTAGATATAAGTGGGT GTGAACAAAAAAGGAAAGGTCGTAATCCTTCACCTACCGC
1921 CCCACGACCCCGCCTACTCCATCTGTCGACCAACTCTCTC AGAAGGTTTATCATGGTCCATCTCATCACTCGGTCCTAC
2001 TTTTCTATCCCGGGTTGGTAAACCCGTTCCCTCTGTCCTA AGGCCCAAATAATACTAATGGGCTTAAATAGTATAAAGTT
2081 CGAATGAAGAAGAAGGAAAAAAAATTGACAAGTTGATTGT TATAGCTACCCATACAGTAGTCAATCGAATGCCAAAGTT
2161 TGGCTTAACTTAACTCATTACCACGTAGCAGAGCATAACA ACAACCACCAATAACATCTAACGGCGTAAACTCACGTGTT
2241 AAGAGAATTACGTCAGAATCCAAGAAAGTGTTTTTTCAG TACACCACAAATAAATTATGTTGAATTTATACCGAACTA
2321 TACTATATTATATATGAAAACATACCTCAATTAGTCATAAT TTCACATTTCATTTAATTTATTAACTAATGGTTGTGTT
2401 GCGGATAGCTTTTCTTTGCTTTCTGTTCTAAAATTATTTC TTAGCCATTATTTGATGATCCAACTTGGTTTGAGGATCTA
2481 ATAAAAATTCAATTTCTTTGAAGCATTCATTGGCTTAATG ATCCAAACCTTTTGTCCTCCACTCTCTAAGAAAATAAAG
2561 CAAAAGCAAAACTAAAAGAAAGCAACTAGTACCGTGGGT TGGTCAATGGTAACCCATGCTACCAAATACCAATCAAAGC
2641 TTAAATGACAATTTAGTACCTTGGGTTGGTCATGATTTAG AGCGGAACAAATTATCCATACATCAAACGAGGATATACAG
2721 AGAAAATTCATTGGAAGTATGGAATTTAGAGGACAATTTCT CTTCTGGGCTACAACGGACGGCCCATTCGCTCATTTACC
2801 CAGAGGTATCGAGTTGTGGACTTTGATGCCGCTAGAGA CTATTGGCATCGGATTGAAAAAAATGTTTACTTCGTTGTT
2881 AACAATTTTCTGAATGCAATATTTCCTTGTCAGAATAT TTAAACTTGTTATTACTTTCTTTGACTTAGGTTGTGGACA
2961 ATTATGGAGTTTACTTCAAACGAGGAAGAATCTTAAACGC TCGGTTCAGGTCTCGAAAACAAACCAACTCACAATCCTGA
3041 CTTAATTGAGGAAAACAATGCAAAACCACATGCATGCTTCC CATTATTTCTATCATAATCTTATAAGAAAAACACTACTAA
3121 GTGAAATGATTCTGTATATATATAACCAATGCCTTTTGTT TTGTGATATTTTATGTATATATAACTATTGACTTTTGTCA
3201 TCTATGGATAGTGTCTCGGGCTCTTGGCAAACATTATTTCA AAGAAAAGTTAATGACTGTAATTAATTAATCTGAAGCTAG
3281 AAACAGAACCCCGAGGTAAAAGAAAAAGACAGAGCACATG AAGTTTAGTACTTTTATATATTTAATATATCATTCTTTCT
3361 TATTGCTTATCTCTAAAGCAAAACTTCCCTAAACCCTAA GCCAAAGGACTCAGATCGATGCAGAACCAAGAAGCCTGT
3441 TTTGGATTTGAGAGCCAAATGCAAAGAAAAAAACTCTTAT GTCTTTCGCAAATGCCGATT
```

Figure 7

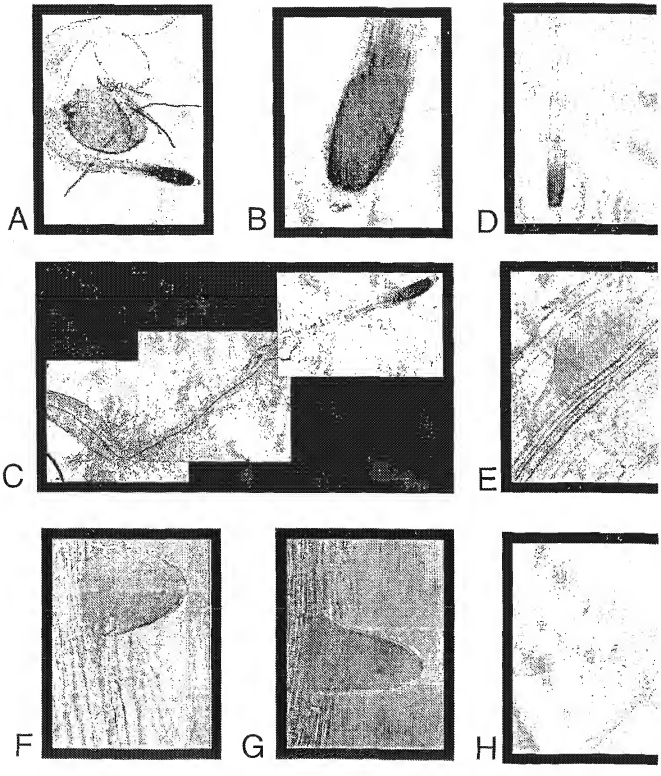


Figure 8 SEQ ID NO 3 : 3960 bp

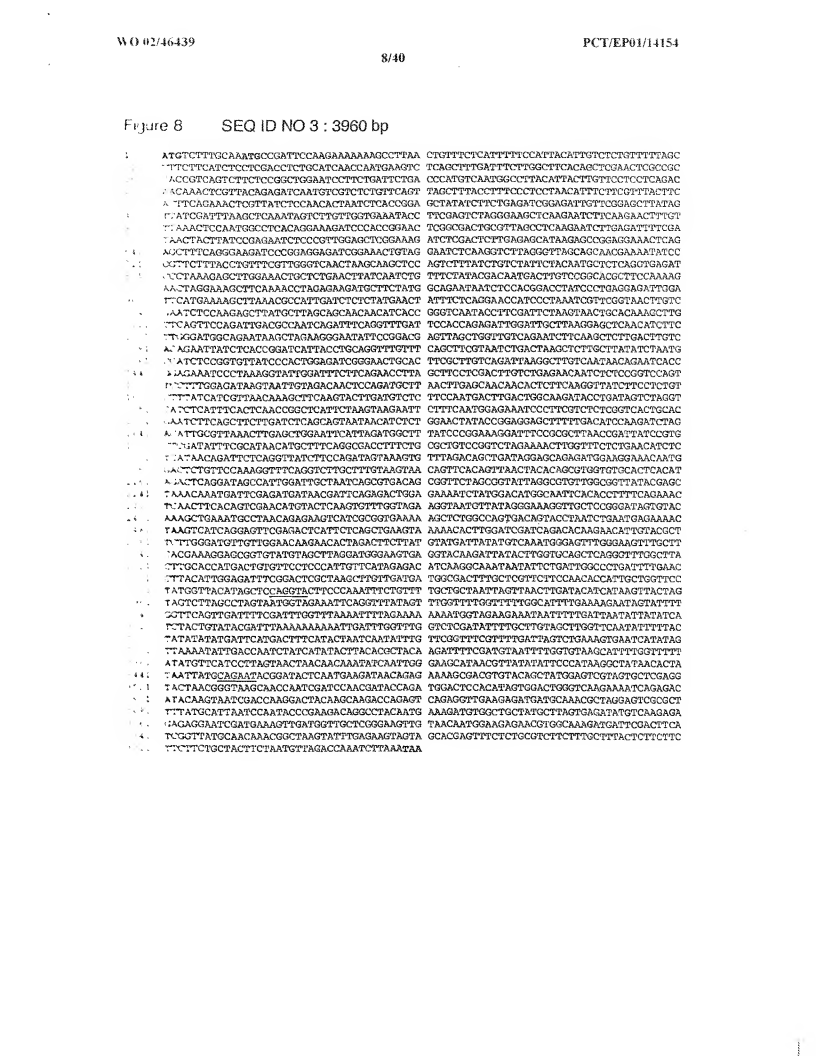


Figure 9

```
1     MSLQMPIPRKKALTVSHPSITLSLFLAPFISSTSASTNEV GALISWLHSSNSPPPSVFSGWNPSDSDPCQWPYITCSSPD
81    NKLVTEINVVSVQLALPFPPNISSFTSLQKLVISNTNLTG AISSEIGDCSELIVIDLSSNSLVGEIPSSLGKLKNLQELC
161   LNSNGLTGKIPPELGDCVSLKNLEIFDNYLSENLPLELGK ISTLESIRAGGNSELSGKIPEEIGNCRNLKVLGLAATKIS
241   GSLPVSLGQLSKLQSLFVYSTMLSGEIPKELGNCSELINL FLYDNDLSGTLPKELGKLQNLEKMLLWQNNLHGPIPEEIG
321   FMKSLNAIDLSMNYFSGTIPKSFGNLSNLQELMLSSNNIT GSIPSILSDCTKLVQFQIDANQISGLIPPEIGLLKELNIF
401   LGWQNKLSGNIPDELAGCQNLQALDLSQNYLTGSLPAGLF QLRNLTKLLLISNAISGVIPLETGNCTSLVRLRLVNNRIT
481   GEIPKGIGFLQNLSFLDLSSNNLSGPVPLEISNCRQLQML NLSWNYLQGYLPLSLSSLTKLQVLDVSSNDLTGKIPDSLG
561   HLISLARLILSKNSFKGEIPSSLGHCTNLQLLDLSGNNIS GTIPEELPDIQDLDIALNLSWNSLDGFIPERISALNRLSV
641   LDISHNMLSGDLSALSGLENLVSLNISHNRFSGYLPDSKV FRQLIGAEMEGNNGLCSKGFRSCFVSNSSQLTTQRGVHSH
721   RLRIAIGLLISVTAVLAVLGVLAVIRAKQMIRDDNDSETG ENLWTWQFTPFQKLNFTVEHVLKCLVEGNVIGKGCSGIVY
801   KAEMPNREVIAVKKLWPVTVPNLNEKTKSSGVRDSFSAEV KTLGSIRHKNIVRFLGCCWNKNTRLLMYDYMSNGSLGSLL
881   HERSGVCSLGWEVRYKIILGAAQGLAYLHHDCVPPIVHRD IKANNILIGPDFEPYIGDFGLAKLVDDGDFARSSNTIAGS
961   YGYIAPEYGYSMKITEKSDVYSYGVVVLEVLTGKQPIDPT IPDGLHIVDWVKKIRDIQVIDQGLQARPESVEEMMQTLG
1041  VALLCINPIPEDRPTMKDVAAMLSRICQEREESMKVDGCS GSCNNGRERGKDDSTSSVMQQTAKYLRSSSTSFSAGSLLY
1121  SSSSSATSHVRPNLK
```

Figure 10-1

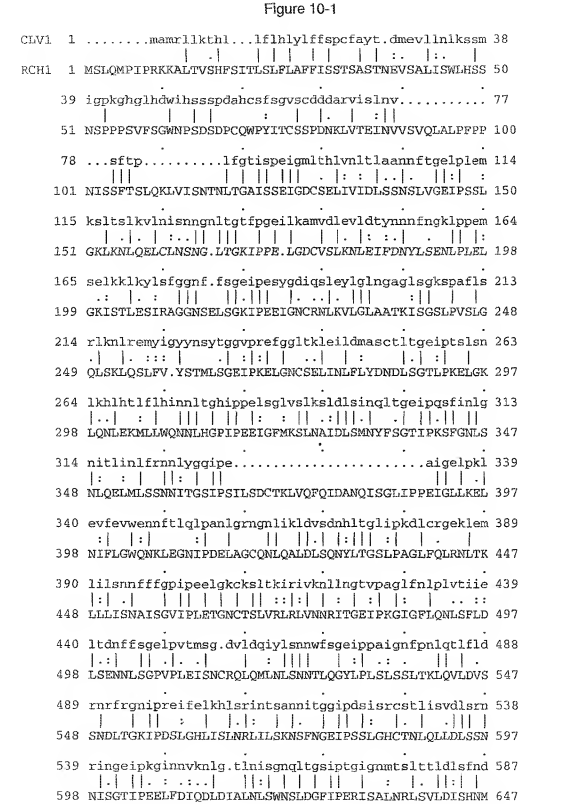

CLV1 1 .....mamrllkthl...lflhlylffspcfayt.dmevllnlkssm 38
RCH1 1 MSLQMPIPRKKALTVSHFSITLSLFLAFFISSTSASTNEVSALISWLHSS 50
39 igpkghglhdwihssspdahcsfsgvscdddarvislnv..... 77
51 NSPPPSVFSGWNPSDSDPCQWPYITCSSPDNKLVTEINVVSVQLALPFPP 100
78 ...sftp.....lfgtispeigmlthlvnltlaannftgelplem 114
101 NISSFTSLQKLVISNTNLTGAISSEIGDCSELIVIDLSSNSLVGEIPSSL 150
115 ksltslkvlnisnngnltgtfpgeilkamvdlevldtynnnfngklppem 164
151 GKLKNLQELCLNSNG.LTGKIPPE.LGDCVSLKNLEIPDNYLSENLPLEL 198
165 selkklkylsfggnf.fsgeipesygdiqsleylglngaglsgkspafls 213
199 GKISTLESIRAGGNSELSGKIPEEIGNCRNLKVLGLAATKISGSLPVSLG 248
214 rlknlremyigyynsytggvprefggltkleildmasctltgeiptslsn 263
249 QLSKLQSLFV.YSTMLSGEIPKELGNCSELINLFLYDNDLSGTLPKELGK 297
264 lkhlhtlflhinnltghippelsglvslksldlsinqltgeipqsfinlg 313
298 LQNLEKMLLWQNNLHGPIPEEIGFMKSLNAIDLSMNYFSGTIPKSFGNLS 347
314 nitlinlfrnnlyggipe.....aigelpkl 339
348 NLQELMLSSNNITGSIPSILSDCTKLVQFQIDANQISGLIPPEIGLLKEL 397
340 evfevwennftlqlpanlgrngnlikldvsdnhltglipkdlcrgeklem 389
398 NIFLGWQNKLEGNIPDELAGCQNLQALDLSQNYLTGSLPAGLFQLRNLTK 447
390 lilsnnfffgpipeelgkcksltkirivknllngtvpaglfnlplvtiie 439
448 LLLISNAISGVIPLETGNCTSLVRLRLVNNRITGEIPKGIGFLQNLSFLD 497
440 ltdnffsgelpvtmsg.dvldqiylsnnwfsgeippaignfpnlqtlfld 488
498 LSENNLSGPVPLEISNCRQLQMLNLSNNTLQGYLPLSLSSLTKLQVLDVS 547
489 rnrfrgnipreifelkhlsrintsannitggipdsisrcstlisvdlsrn 538
548 SNDLTGKIPDSLGHLISLNRLILSKNSFNGEIPSSLGHCTNLQLLDLSSN 597
539 ringeipkginnvknlg.tlnisgnqltgsiptgignmtslttldlsfnd 587
598 NISGTIPEELFDIQDLDIALNLSWNSLDGPIPERISALNRLSVLDISHNM 647

Figure 10-2

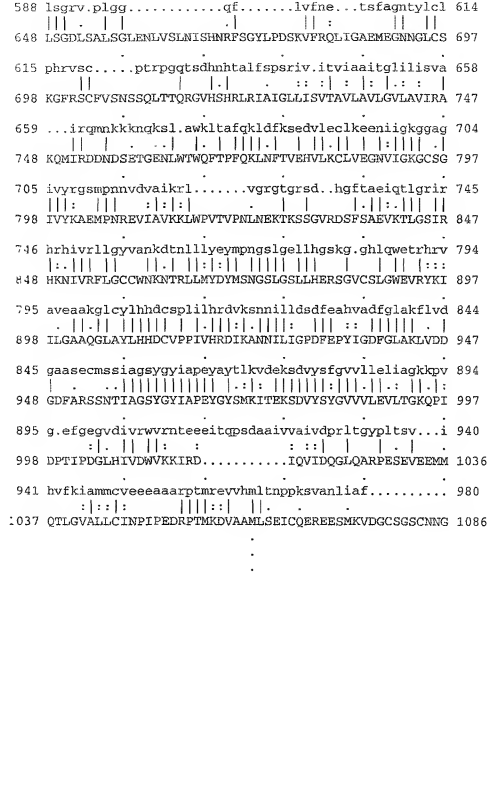

588 lsgrv.plgg.....qf.....lvfne...tsfagntylcl 614
648 LSGDLSALSGLENLVSLNISHNRFSGYLPDSKVFRQLIGAEMEGNNGLCS 697
615 phrvsc.....ptrpgqtsdhnhtalfspsriv.itviaaitglilisva 658
698 KGFRSCFVSNSSQLTTQRGVHSHRLRIAIGLLISVTAVLAVLGVLAVIRA 747
659 ...irqmnkkknqksl.awkltafqkldfksedvleclkeeniigkggag 704
748 KQMIRDDNDSETGENLWTWQFTPFQKLNFTVEHVLKCLVEGNVIGKGCSG 797
705 ivyrgsmpnnvdvaikrl.....vgrgtgrsd..hgftaeiqtlgrir 745
798 IVYKAEMPNREVIAVKKLWPVTVPNLNEKTKSSGVRDSFSAEVKTLGSIR 847
746 hrhivrllgyvankdtnlllyeympngslgellhgskg.ghlqwetrhrv 794
848 HKNIVRFLGCCWNKNTRLLMYDYMSNGSLGSLLHERSGVCSLGWEVRYKI 897
795 aveaakglcylhhdcsplilhrdvksnnilldsdfeahvadfglakflvd 844
898 ILGAAQGLAYLHHDCVPPIVHRDIKANNILIGPDFEPYIGDFGLAKLVDD 947
845 gaasecmssiagsygyiapeyaytlkvdeksdvysfgvvlleliagkkpv 894
948 GDFARSSNTIAGSYGYIAPEYGYSMKITEKSDVYSYGVVVLEVLTGKQPI 997
895 g.efgegvdivrwvrnteeeitqpsdaaivvaivdprltgypltsv...i 940
998 DPTIPDGLHIVDWVKKIRD.....IQVIDQGLQARPESEVEEMM 1036
941 hvfkiammcveeeaaarptmrevvhmltnppksvanliaf..... 980
1037 QTLGVALLCINPIPEDRPTMKDVAAMLSEICQEREESMKVDGCSGSCNNG 1086

Figure 11

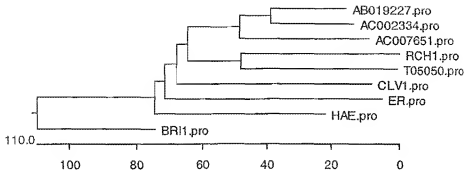

AB019227.pro
AC002334.pro
AC007651.pro
RCH1.pro
T05050.pro
CLV1.pro
ER.pro
HAE.pro
BRI1.pro
110.0
100
80
60
40
20
0

Figure 12

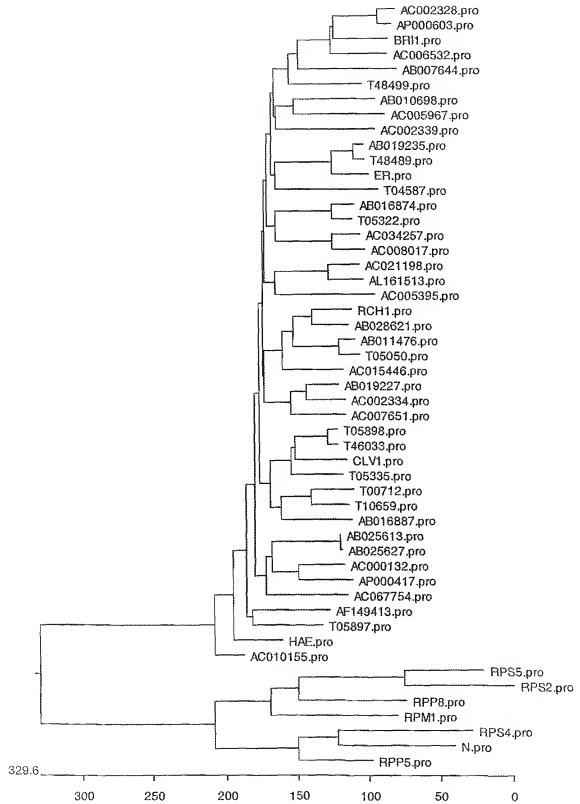
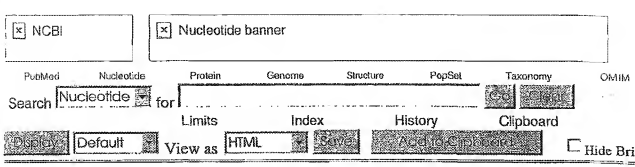

AC002328.pro
AP000603.pro
BRI1.pro
AC006532.pro
AB007644.pro
T48499.pro
AB010698.pro
AC005967.pro
AC002339.pro
AB019235.pro
T48489.pro
ER.pro
T04587.pro
AB016874.pro
T05322.pro
AC034257.pro
AC008017.pro
AC021198.pro
AL161513.pro
AC005395.pro
RCH1.pro
AB028621.pro
AB011476.pro
T05050.pro
AC015446.pro
AB019227.pro
AC002334.pro
AC007651.pro
T05898.pro
T46033.pro
CLV1.pro
T05335.pro
T00712.pro
T10659.pro
AB016887.pro
AB025613.pro
AB025627.pro
AC000132.pro
AP000417.pro
AC067754.pro
AF149413.pro
T05897.pro
HAE.pro
AC010155.pro
RPS5.pro
RPS2.pro
RPP8.pro
RPM1.pro
RPS4.pro
N.pro
RPP5.pro
329.6
300
250
200
150
100
50
0

Figure 13 SEQ ID NO 18 : 842 bp

AAGCTTAAATGACAATTTAGTACCTTGGGTTGGTCATGATTTAGAGCGGAACAAATATACCAT
ACATCAAACGAGGATATACAGAGAAAATTCATGGAAGTATGGAATTTAGAGGACAATTTCTC
TTCTGGGCTACAACGGACCGGCCCATTCGCTCATTTACCCAGAGGTATCGAGTTTGTGGACTTT
TGATGCCGCTAGAGACTATTGGCATCGGATTGAAAAAAATGTTTACTTCGTTGTTAACAATTTT
CTGAATGCAATATTTTCCTTGTCATGAATATTTAAACTTGTTATTACTTTCTTTTAGCTTAGGTG
TGGACAATTATGGAGTTTACTTCAAACGAGGAAGAATCTTAAACGCTCGGTTCAGGTCTCGAA
AACAAACCAACTCACAATCCTGACTTAATTGAGGAAAACAATGCAAAACCACATGCATGCTT
CCATATTTCTATCATAATCTTATAAGAAAAAACACTACTAAGTGAAATGATTCTGTATATATAT
AACCAATGCCTTTTGTTTTGTGATATTTTATGTATATATAACTATTGACTTTTGTCATCTATGGA
TAGTGTCTCGGGCTCTTGGCAAACATATTTCAAAGAAAAGTTAATGACTGTAATTAATTAATC
TGAAGCTAGAAACAGAACCCCGAGGTAAAAGAAAAAGACAGAGCACATGAAGTTTAGTACTT
TTATATATTTAATATATCATTCTTTCTTATTGCTTATCTCTAAAGCAAAAACTTCCCTAAACCCT
AAGCCAAAGGACTCAGATCGATGCAGAACCAAGAAGGCTTGTTTTGGATTTGAGAGCCAAAT
GCAAAGAAAAAAACTCTT

NCBI Sequence Viewer

Figure 14




```
LOCUS       AB017061     61712 bp    DNA              PLN       09-AUG-2000
DEFINITION  Arabidopsis thaliana genomic DNA, chromosome 5, TAC clone:K19E20.
ACCESSION   AB017061
VERSION     AB017061.1  GI:3510337
KEYWORDS    .
SOURCE      Arabidopsis thaliana (strain:Columbia) DNA, clone_lib:Mitsui TAC
            clone:K19E20.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
            Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (sites)
  AUTHORS   Kaneko,T., Katoh,T., Sato,S., Nakamura,Y., Asamizu,E., Kotani,H.,
            Miyajima,N. and Tabata,S.
  TITLE     Structural analysis of Arabidopsis thaliana chromosome 5. IX.
            Sequence features of the regions of 1,011,550 bp covered by
            seventeen P1 and TAC clones
  JOURNAL   DNA Res. 6 (3), 183-195 (1999)
  MEDLINE   99397451
REFERENCE   2  (bases 1 to 61712)
  AUTHORS   Nakamura,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-AUG-1998) to the DDBJ/EMBL/GenBank databases.
            Yasukazu Nakamura, Kazusa DNA Research Institute, Department of
            Plant Gene Research; 1532-3, Yana, Kisarazu, Chiba 292-0812, Japan
            (E-mail:ynakamu@kazusa.or.jp, Tel:81-438-52-3935,
            Fax:81-438-52-3934)
COMMENT     Address for correspondence: kaos@kazusa.or.jp
            For the latest information on annotation of this clone, please see
            http://www.kazusa.or.jp/kaos/cgi-bin/agd_graph.cgi?c=K19E20
            Genes with similarity to proteins in the databases are described in
            'product' or 'note' qualifiers.  Genes that have no significant
            protein similarity are described as 'unknown protein'.
            The software programs used to predict genes include: Grail
            (Informatics Group, Oak Ridge National Laboratory,
            http://compbio.ornl.gov/Grail-1.3/),
            GENSCAN (Chris Burge, MIT, http://CCR-081.mit.edu/GENSCAN.html),
            NetGene2 (S.M. Hebsgaard, et al., CBS, Technical University of
            Denmark, http://www.cbs.dtu.dk/services/NetGene2/) and
            SplicePredictor (Volker Brendel, Stanford University,
            http://gremlin1.zool.iastate.edu/cgi-bin/sp.cgi).
            Genes encoding tRNAs are predicted by tRNAscan-SE
            (Sean Eddy, Washington University School of Medicine, St. Louis,
            http://genome.wustl.edu/eddy/tRNAscan-SE/).
            This sequence may not be the entire insert of this clone. It may be
            shorter because we remove overlaps between neighboring submissions.
            The 5' clone is K24G6 and the 3' clone is K20J1.
FEATURES            Location/Qualifiers
```

NCBI Sequence Viewer

source 1..61712
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/chromosome="5"
/clone="K19E20"
/clone_lib="Mitsui TAC"

CDS join(1745..1821,3560..4202,4287..4390,4470..4634,
4781..4973)
/note="gene_id:K19E20.1"
/codon_start=1
/evidence=not_experimental
/product="pectate lyase"
/protein_id="BAB10313.1"
/db_xref="GI:10177179"
/translation="MTSRSNETWHEHAVENPEEVAAMVDMSIRNSTARRRLGYFSCST
GNPIDDCWRCDRRWQSRRKHLANCAIGFGRNAIGGRDGRYYVVSDPNDDNPVNPKPGT
LRHAVIQEEPLWIVFKRDMVITLKEELIMNSFKTIDGRGVNVHIANGACITIQFVTNI
IIHGIHIHDCRPTGNAMVRSSPSHYGWRTMADGDGISIFGSSHIWIDHNSLSNCADGL
IDAVMASTAITISNNYFTHHNEVMLLGHSDTYTRDKVMQVTIAYNHFGEGLIQRMPRC
RHGYFHVVNNDYTHWEMYAIGGSASPTINSQGNRYLAPRNRFAKEVLSRVRSTRQWRH
WNWRSEGDLFLNGAFFTRSGSGLGASYARASSLAAKSSSLVGVITYNAGALNCRGGRR
C"

CDS complement(9033..10973)
/note="gene_id:K19E20.2"
/codon_start=1
/evidence=not_experimental
/product="selenium-binding protein-like"
/protein_id="BAB10314.1"
/db_xref="GI:10177180"
/translation="MNPTQTLFSPGGNSPASSPASHPSSLFPQINNCRTIRDLSQIHA
VFIKSGQMRDTLAAAEILRFCATSDLHHRDLDYAHKIFNQMPQRNCPSWNTIIRGFSE
SDEDKALIAITLFYEMMSDEFVEPNRFTFPSVLKACAKTGKIQEGKQIHGLALKYGFG
GDEFVMSNLVRMYVMCGFMKDARVLFYKNIIEKDMVVMTDRRKRDGEIVLWNVMIDGY
MRLGDCKAARMLFDKMRQRSVVSWNTMISGYSLNGFFKDAVEVFREMKKGDIRPNYVT
LVSVLPAISRLGSLELGEWLHLYAEDSGIRIDDVLGSALIDMYSKCGIIEKATHVFER
LPRENVITWSAMINGFAIHGQAGDAIDCFCKMRQAGVRPSDVAYINLLTACSHGGLVE
EGRRYFSQMVSVDGLEPRIEHYGCMVDLLGRSGLLDEAEEFILNMPIKPDDVIWKALL
GACRMQGNVEMGKRVANILMDMVPHDSGAYVALSNMYASQGNWSEVSEMRLRMKEKDI
RKDPGCSLIDIDGVLHEFVVEDDSHPKAKEINSMLVEISDKLRLAGYRPITTQVLLNL
EEEDKENVLHYHSEKIATAFGLISTSPGKPIRIVKNLRICEDCHSSIKLISKVYKRKI
TVRDRKRFHHFQDGSCSCMDYW"

CDS 11469..12086
/note="gene_id:K19E20.3
unknown protein"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB10315.1"
/db_xref="GI:10177181"
/translation="MAASVEYFPYYSPPSHQHPLPSPVPPPPSHISPPPPFSPPHHP
PPPHFSPPHQPPPSPYPHPHPPPPSPYPHPHQPPPPPHVLPPPPPTPAPGHHVIIVVV
ISLGSLFFLAFLAAALFCYLKKRRKSSTKAEIIEFDEHLKVQETIVQGPHGEQTRVVM
LEEDIHLVEDIHKTEKLSRPSHLSSTGRHAIDISDPNHHFTEQKS"

CDS complement(join(12718..13611,13749..14156))
/note="gene_id:K19E20.4"
/codon_start=1
/evidence=not_experimental
/product="anthranilate N-benzoyltransferase"
/protein_id="BAB10316.1"
/db_xref="GI:10177182"
/translation="MKINIRDSTMVRPATETPITTNLWNSNVDLVIPRFHTPSVYFYRP
TGASNFFDPQVMKEALSKALVPFYPMAGRLKRDDDGRIEIDCNGAGVLFVVADTPSVI
DDFGDFAPTLNRLLIPEVDHSAGIHSFPLLVLQVTFFKCGGASLGVGMQHHAADGFS
GLHFINTWSDMARGLDLTIPPFIDRTLLRARDPPQPAFHHVEYQPAPSMKIPLDPSKS

NCBI Sequence Viewer

GPENTTVSIFKLTRDQLVALKAKSKEDGNTVSYSSYEMLAGHVWRSVGKARGLPNDQE
TKLYIATDGRSRLRPQLPPGYFGNVIFTATPLAVAGDLLSKPTWYAAGQIHDFLVMGD
DNYLRSALDYLEMQPDLSALVRGAHTYKCPNLGITSWVRLPIYDADFGWGRPIFMGPG
GIPYEGLSFVLPSPTNDGSLSVAIALQSEHMKLPEKFLFEI"

CDS join(15849..18529,18605..18747,19300..19808)
/note="gene_id:K19E20.5"
/codon_start=1
/evidence=not_experimental
/product="receptor protein kinase-like protein"
/protein_id="BAB10317.1"
/db_xref="GI:10177183"
/translation="MSLQMPIPRKKALTVSHPSITLSLFLAFFISSTSASTNEVSALI
SWLHSSNSPPPSVFSGWNPSDSDPCQWPYITCSSSDNKLVTEINVVSVQLALPFPPNI
SSFTSLQKLVISNTNLTGAISSEIGDCSELIVIDLSSNSLVGEIPSSLGKLKNLQELC
LNSNGLTGKIPPELGDCVSLKNLEIFDNYLSENLPLELGKISTLESIRAGGNSELSGK
IPEEIGNCRNLKVLGLAATKISGSLPVSLGQLSKLQSLSVYSTMLSGEIPKELGNCSE
LINLFLYDNDLSGTLPKELGKLQNLEKMLLWQNNLHGPIPEEIGFMKSLNAIDLSMNY
FSGTIPKSFGNLSNLQELMLSSNNITGSIPSILSNCTKLVQFQIDANQISGLIPPEIG
LLKELNIFLGWQNKLEGNIPDELAGCQNLQALDLSQNYLTGSLPAGLFQLRNLTKLLL
ISNAISGVIPLEIGNCTSLVRLRLVNNRITGEIPKGIGFLQNLSFLDLSENNLSGPVP
LEISNCRQLQMLNLSNNTLQGYLPLSLSSLTKLQVLDVSSNDLTGKIPDSLGHLISLN
RLILSKNSFNGEIPSSLGHCTNLQLLDLSSNNISGTIPEELFDIQDLDIALNLSWNSL
DGFIPERISALNRLSVLDISHNMLSGDLSALSGLENLVSLNISHNRFSGYLPDSKVFR
QLIGAEMEGNNGLCSKGFRSCFVSNSSQLTTQRGVHSHRLRIAIGLLISVTAVLAVLG
VLAVIRAKQMIRDDNDSETGENLWIWQFTPFQKLNFTVEHVLKCLVEGNVIGKGCSGI
VYKAEMPNREVIAVKKLWPVTPNLNEKTKSSGVRDSFSAEVKTLGSIRHKNIVRFLG
CCWNKNTRLLMYDYMSNGSLGSLLHERSGVCSLGWEVRDIKANNILIGPDFEPYIGDF
GLAKLVDDGDFARSSNTIAGSYGYIAPEYGYSMKITEKSDVYSYGVVVLEVLTGKQPI
DPTIPDGLHIVDWVKKIRDIQVIDQGLQARPESEVEEMMQTLGVALLCINPIPEDRPT
MKDVAAMLSEICQEREESMKVDGCSGSCNNGRERGKDDSTSSVMQQTAKYLRSSSTSF
SASSLLYSSSSSATSNVRPNLK"

CDS join(22264..22578,23196..23354)
/note="gb|AAD49765.1
gene_id:K19E20.6
similar to unknown protein"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB10318.1"
/db_xref="GI:10177184"
/translation="MDPKSPEFIIDQPLKILGFVFDELSATRVSGHLTLTEKCCQPFK
VLHGGVSALIAEALASLGAGIASGFKRVAGIHLSIHHLRPAALGEIVFAESFPVSVGK
NIQVWEVRLWKAKKTETPDNKIMVSTSRVTLFCGLPIPDHVKDAPDELKKVISKL"

CDS join(25778..26072,26283..26377,26531..26650,26729..26818,
26916..27031,27145..27244,27447..27518,27597..27644,
27726..27803,27879..27998,28105..28210,28303..28466,
28574..28708,28845..28975,29226..29304,29399..29446)
/note="contains similarity to 5'-nucleotidase
gene_id:K19E20.8"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB10319.1"
/db_xref="GI:10177185"
/translation="MNLCADQVFDKKLLWRDMSTKMKFPSFSAAELPDLRKSNKRRGS
LRMIKCRAAGADGGRVAVGDDVFSVTTSSKYEVDYLGQSTKGDLNLKLDPLQSFGDGQ
ATLEGPIEEVARTEAQAAENLIRELGIQGPFSAQHSPRGIFCSRTLNLRSISAIGYDM
DYTLMHYNVMAWEGKAYDYCMENLKSMGFPVDGLAFDPELVIRGLMIDKEKGNLVKAD
RPGYVKRAMHGTKMLSNKAVSEIYGRELVDLRNQSRWEFLNTFFSVSEALAYAQMVDR
LDDGFISADLGTLDYKGLYKAVAKALPRAHVEGQLKSEIMSKPELFVEPDPELPLALL
DQKEAGKKLLLITNSDYHYTDKMMKHSFNKFLPNDMDWRDLFDMVIVSARKPEFFQMS
HPLYEVVTGEGLMRPCFKAETGGLYSGGSAQMIESSLNVHGDELILYVGDHIYTDVSVS
KVHLRWRTALICRELEEEYMALIGSRGHREELIELINQKEVVGDLFNQLRLALQRRSK
GRPAQTLAATNLDDQELTETMQKLIVMQRLDDKIGLMLETDGELFNKRYADIYTSRV
SNFLNYTPFMYFRSQEQSLAHDSPLPDAGIEN"

CDS join(30274..30620,30707..32633)

NCBI Sequence Viewer

/note="gene_id:K19E20.9"
/codon_start=1
/evidence=not_experimental
/product="mutator-like transposase-like protein"
/protein_id="BAB10320.1"
/db_xref="GI:10177186"
/translation="MATKKVIAICQSGGEFVTNKDGSLAYSGGDAYAIDIDQDTCMSD
FKSELAENFGFSWENMTLKYFLPGNKKTLITISKDKDFQRMVSFSADAPNVEIFVLPE
EAEARNVSNMPASRSSRTTASEAVVPVVSAGDIVVGDDMIANDFQIEMGIMPPESSPL
PCNFVLTDEKQHIKAAQQWENAITGVDQRPNSFTEFRDALHKYSIAHGFTYKYKKNDS
HRVSVKCKAQGCPWRITASRLSTTQLICIKKMNTRHTCERAVVKAGYRASRGWVGSII
KEKLKAFPDYKPKDIAEDIKREYGIQLNYSQAWRAKEIARSQLQGSYKKAYSQLPSFC
KKIRETNPGSIAIFMTKEDSSFHRLFISFYASISGFRQGCRFLLFLDTADLNSKYQGV
MLVATAPDAEDGIFPVAFAVVDAETEDNMVWFLEHLKLALADPRTITFVADFQNGLKT
ALPLVFEKQHHHAYCLRHLAEKLNMDLQAQFSHEARRFILNDFYAAAYATQPDAYYRS
LENIKSISPDAYTWVIESEPLHWANALFEGERYNHMNSIFGLDFYSWVSEAHELPITQ
MIDELRAKLMQSIYTHQVQSREWIVSTLTPTNEEKLQKEIELARSLQVSAPHNSLPEV
HGETINLVDINQCDCDCKVWRLTGLPCSHAVAVVECIGKSPYEYCSRYFTSESYRLTY
AESINPVPNTTMTMMILEEPFVEGVVSVTPPPTRLTPPGRPKSKQVEPLDMFKRQLQC
SNCKGLGHNKKTCKAVS"
CDS complement(join(33092..33310,33890..34039,34118..34270,
34394..34521,34741..34852,34989..35105,35298..35438))
/note="gene_id:K19E20.10"
/codon_start=1
/evidence=not_experimental
/product="mitochondrial carrier protein-like"
/protein_id="BAB10321.1"
/db_xref="GI:10177187"
/translation="MIPTTATVDDEPGQIKRALIDASAGAISGGVSRSVTSPLDVIKI
RFQVQLEPTTSWGLVRGNLSGASKYTGMVQATKDIFREEGFRGFWRGNVPALLMVMPY
TSIQFTVLHKLKSFASGSTKTEDHIHLSPYLSFVSGALAGCAATLGSYPFDLLRTILA
SQGEPKVYPTMRSAFVDIIQSRGIRGLYNGLTPTLVEIVPYAGLQFGTYDMFKRWMMD
WNRYKLSSKIPINVDTNLSSFQLFICGLGAGTSAKLVCHPLDVVKKRFQIEGLQRHPR
YGARVERRAYRNMLDGLRQIMISEGWHGLYKGIVPSTVKAAPAGAVTFVAYEFTSDWL
ESISW"
CDS 36788..37987
/note="gene_id:K19E20.11
pir||T09376
strong similarity to unknown protein"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB10322.1"
/db_xref="GI:10177188"
/translation="MADSQRLSTASGVKDGQPPWKKKKLSNDTTSNPSLPYDVILIIL
ARVSRSYYTNLSLVSKSFSRSILTSPELYKTPRTLLGKTENFLYVCLRFPDEANPRWFTL
YRKPNQTLTDHTTKKKKKKKKKEEKSSVNLLAPISILNSHPVEWSAIISVDHYLYAIS
ADIEKAPYSNVPYLDCRTHTWNEAPRMRLAHTNSEFEGIVYLPGSFESPDSLNCVEVY
NTMTQTWKPVPPEKRMFKLENLEKKIYYKSFHLDSRAGKGLSLSYKSKHLTCGLVVLD
TVDSYLRSSCMIENIAYFYRKGNPIWRGLDGKILVYGKIEGLEGLPKFSRYSSVQLAE
YGGKLVVLWDKYVPASGYKEKMIWCAEISLEKRNGKEIWGNVEWPDAVLTVPKSYKIL
CATAATL"
CDS 38404..39522
/note="gene_id:K19E20.12
pir||T09376
strong similarity to unknown protein"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB10323.1"
/db_xref="GI:10177189"
/translation="MSSPEKKRKKNQKSSPNPSLPEDLIVSILARVSRSYYTNLSVVS
KTPRSILTSPELYKTRTLLGKPETFLYVCLRFPDEANPRWLILYRKPNQTLTKKKKKK
EDSSVHLLAPIPILNSPTVEWSSLVAVGSYLYATTADIKDSPCSNVWYLDCRTHTWLD
SPRLRLAHINSDPNGRTYFPGSSSEKPDSLNCVEVYNTNTQTWNPVPPQKRKLKFGNME
GKIYIPPCQENHRKAVALNPKVLTWEAVGLGTNLDRGSFCMIGNIAYCYDPSGKFRWM

NCBI Sequence Viewer

```
                NCNTAEGDWSRLEGLEGLPRFARYSTVKLAEYGGKLVVLWDKYVAASGYKEKMIWCAE
                ISLEKRNSEEIWGKVEWFDAVLTVPKSYKFVCAKSATV"
CDS             40591..41709
                /note="gene_id:K19E20.13
                pix||T09371
                strong similarity to unknown protein"
                /codon_start=1
                /evidence=not_experimental
                /protein_id="BAB10324.1"
                /db_xref="GI:10177190"
                /translation="MSSPERKRKKRSLEPSPESTPNPSLPDDLIVSILARVSRLYYPI
                LSLVSKSSRTLVTSPELYKTRSFPNRTESCLYVCLDFPPDPNPRWFTLYRKPNQNITE
                KTKNSSGFVLAPIPNHHSHSSSIVAIGSNIYAIGGSIENAPSSKVSILDCRSHTWHEA
                PSMRMKRNYPAANVVDGKIYVAGGLEEFDSSKWMEVFDIKTQTWEFVLSPLAERFIYR
                SLVIEGEIYIFGDKVVTYKPKEDRWGGVGEHQSMDLGLFFHSYCVIDNVLYCYRPGGI
                KWYESEKRSWRKLRGLKGLSKLASSCVKLADYGGKMALLWDKYIPCSGNKSHSISCAV
                VSLERCKNQGIRGKVEWFDDMLTVPSSYNFVGALAATL"
CDS             join(42528..42695,42932..43054,43183..43265,43625..43734,
                43820..44100)
                /note="gene_id:K19E20.14
                unknown protein"
                /codon_start=1
                /evidence=not_experimental
                /protein_id="BAB10325.1"
                /db_xref="GI:10177191"
                /translation="MASNSEAGGSADYETLMSTSDVELLKRAWRNEKAAPEILQYEGA
                LVDRAKEQIELVEETIEDYVENGIDPLVVSLYQMDLDRAQFLLRSYLRVRLLKIEKFM
                FHNLNSEEAERRLSEQEKVFATRCADDLAKHFEETVLLKLPENYQSVLKQSLISEVDD
                MVARVVFVLVMTTLVNLQCRSRIWIPLLFVEARTLSPSIYTKSEHSLIPPYTYIYTQS
                FCNLSVGLPLLTLAEERAPRLWRWSAVISTSYVTRL"
CDS             join(47405..47596,47798..47863,48017..48125,48257..48403,
                48519..48622,48732..48830,48921..48990,49159..49255,
                49331..49385,49512..49595,49681..50009,50129..50301,
                50385..50521,50676..50747)
                /note="gene_id:K19E20.17"
                /codon_start=1
                /evidence=not_experimental
                /product="arginine methyltransferase-like protein"
                /protein_id="BAB10326.1"
                /db_xref="GI:10177192"
                /translation="MEIPSLNKQQEFTLASVTDLTSPSSSLSSSPVVATFSCVNEVKE
                LRFQESKSSDGFSFDLSSTQLFKLGPLQFTCVSDGSISSAKEKSSFSRGVVIKFRDEK
                DSKEFCDSFEECKKDDAVKQGSALPNGTVVSANKSKFDDKIEAASAKMYFHYYGQLLH
                QQNMLQDYVRTGTYHAAVMENRSDFSGRVVVDVGAGSGILSMFAALAGAKHVYAVEAS
                EMAEYARKLIAGNPLLAERITVIKGKIEDIELPEKADVLISEPMGTLLVNERMLETYV
                IARDRFLSPNGKMFPTVGRIHMAPFADEFLFVEMANKALFWQQQNYYGVDLTPLYVSA
                HQGYFSQPVVDAFDPRLLVAPSMFHVIDFTMMTVKLFPTLDSVPRCFLKIEDTFLSCY
                FYVQSSLFSEWHDPNSITVKFFSQEEQFYEIDIPLKFTASVCTRIHGLACWFDVLFDG
                STVQRWFTTAPGAPTTHWYQIRCVLSQPIHVMAGQEITGRLHLIAHSAQSYTINLTLS
                AKMWGPGANQGGILQTSSCKLDLKEPYYRMSQPQVYPTQEPPAQSQDIHIHSDDLEEL
                ELLQQNANAQL"
CDS             complement(join(51155..51406,51952..52131,52447..52696,
                52806..52879,53145..53270,53408..53572,53658..53785,
                53900..54028,54123..54254,54347..54494,54771..54907,
                54997..55283,55380..55498,55559..55611,55791..55935,
                56245..56352,56467..56535,56693..56805,56925..56994,
                57106..57207,57282..57436,57516..57540))
                /note="gene_id:K19E20.18"
                /codon_start=1
                /evidence=not_experimental
                /product="isoleucyl-tRNA synthetase"
                /protein_id="BAB10327.1"
                /db_xref="GI:10177193"
                /translation="MAAKKASEGEKQEDGKYKQTVDLPKTGFGMRANSLTREPELQKL
```

NCBI Sequence Viewer

```
                WEENQVFKRVSDNNNGGSFILHDGPPYANGDLHMGHALNKILKDIINRYKLLQNYKVQ
                YVPGWDCHGLPIELKVLQSLDQEVRKELTPLKLRAKAAKFAKATVKTQMESFKRFGVW
                ADWNNPYLTLDPEYEAAQIEVFGQMALKGYIYRGRKPVHWSPSSRTALAEAELEYPEG
                HISKSIYAIFKLVGGAKTSLLDEFIPNTYLAVWTTTPWTMPANAAVAVNAKLQYSVVE
                VQSFNQQKLFVIVATDLVPALEAKWGVKLSISKTFLGSDLENCRYTHPIDNRDCPVVI
                GGDYITTESGTGLVHTAPGHGQEDYATGLKYGLPLVSPVDDEGKFTEEAGQFRGLSVL
                GEGNTAVVSYLDENNMSLVMEESYAHKYPYDWRTKKPTIFRATEQWFASVEGFRTATMD
                AINNVKWVPHQAVNRISAMTSSRSDWCISRQRTWGVPIPAFYHVKTLKPEMNEETINH
                VKSIISQKGSDAWWYMSVEDLLPEKYRDKAADYEKGTDTMDVWFDSGSSWAGVLGKRE
                GLSFPADVYLEGTDQHRGWFQSSLLTSIATQGKAPYSAVITHGFVLDEKGMKMSKSLG
                NVVDPRLVIEGGKNSKDAPAYGADVMRLWVSSVDYTGDVLIGPQILRQMSDIYRKLRG
                TLRYLLGNLHDWRVDNAVPYQDLPIIDQHALFQLENVVKNIQECYENYQFFKIFQIIQ
                RFTIVDLSNFYFDIAKDRLYTGGTSSFTRRSCQTVLSTHLLSILRVIAPIVPHLAEDV
                WQNLPFEYRNEDGSAAEFVFELKWFTLNEQWLSFPAEDVLFPWQRLLELRTEVNKVLEL
                ARNEKMIGSSLEAKVYLHTADAGMAAKLLEMSEAKNEADTLQRIFITSQVEVLSSMEK
                EMISSVQHTGEYVEGENKVWIGVSRAEGSKCERCWNYSGQVGSPSDHPTLCGRCFSVI
                VANPPEPAVAAVNSLA"
     CDS        complement(58740..59315)
                /note="contains similarity to disease resistance response
                protein
                gene_id:K19E20.19"
                /codon_start=1
                /evidence=not_experimental
                /protein_id="BAB10328.1"
                /db_xref="GI:10177194"
                /translation="MSKLILILTAQILLLTATALAGKNGEDFARTINRKHLGLGKKEK
                LTHLRVYWHDIVTGRNPSSIRIQGPVAKYSSSSYFGSITMIDNALTLDVPINSTVVGQ
                AQGMYVGAAQKEIGLLMAMNLAFKTGKYNGSTITILGRNTVMSKVREMPVVGGSGMFR
                FARGYVEARTKLFDMKTGDATVESNCYILHY"
     CDS        59993..60445
                /note="gene_id:K19E20.20
                unknown protein"
                /codon_start=1
                /evidence=not_experimental
                /protein_id="BAB10329.1"
                /db_xref="GI:10177195"
                /translation="MPRVNIGPDNNAQIEITTLQSQSKSRKLKKIKMASGSERVVVVG
                VDDSAHSYHALETALDLFFIPPKANPRPKLVVLHARPTATFFLGVAGPGTVDIIPMVE
                EDLNKTADLVKKKCAEVCSAKSVEISSLEVIEGDPRNIMLEAVERHHA"
     misc_RNA   join(complement(AB023028.1:1..96),
                complement(61652..61712))
                /note="gene_id:K19E20.21"
                /product="U4 snRNA"
                /evidence=not_experimental
BASE COUNT      19394 a  11344 c  11242 g  19732 t
ORIGIN
        1 tgatcaacat gactaatctt ccacgagact agaaccacta acccacctcc tctagataaa
       61 gctaaacaat tctctacaaa aaatttccat taaatgtttc taattgtaat tgaaaacatc
      121 aacttaaaat attattaaat aaattttgta taaacgtgat tcatttaac ttttttttta
      181 atgaattcta aatccaaatg tgtcttaatc tactgctatc acaactccac aatattagaa
      241 acaaaacata tagagcttgg gccattttct gcttttatga ttttcacgaa gtgttttcac
      301 caatatactt acatcactag ttcagttgaa tctacttaaa atcataaggt acgtaactca
      361 aatcacatta acgattcaag ttcattttca tggaattaga tattttagta aaattactac
      421 ctattgatga atttggatat ttctagtttt aaaacgggag gaaaaaatga ttaacgtctt
      481 tcatagcact atataaatgt taatttttat ttgaattaaa aatattcagt taacgaaagt
      541 gtttgtgatt aatctagagt ttatttggag aaatagtcaa aaaaaagttt ataatacctt
      601 tttcctggtt catttttttt ttttttgcc aacatttaaa ttccaattgt tgagatttta
      661 gcccatcatt taatggtgag aagaatgaaa gtgctaaact atttgtctta aatgattata
      721 ctttgattat aaaactgtga caatatgatt attatacata tgtttaaatt ttatttctatg
      781 tgatacatat ataaatgttt tagcaagtac aaaacatgga aaatgtgagc gacttacaat
      841 tataacatca attctcatta gcatttccgc tttactgaa aacaaaaacg agtgagtaca
      901 atttgtcttc ccttttttt ttttttggac acaaagtaca atttatcttc tatttcctcc
      961 tttttgtgta atttgtctct tgtcgtatca cctacatttt cagcattgga tgttataaat
     1021 gcgaacaaac tataatgtaa aaaacatgtc agcagcatcg ggccatttca aactatgtaa
```

NCBI Sequence Viewer

```
1081 ataatgaaga aattatatga tctattctaa ataattaatt aggccaaata ttatttctaag
1141 tgaaaaatgt cacgctcaat ttctcttata taataactcc ttccttgaa gccttttctt
1201 aactcttttc atcttccttc ttcttcttc ttctttaaac acaaaaaacc gacaaaagcc
1261 tctgtctcgg acagagacaa agaaacagaa cagagcaaga acaatggctg ttacacaaat
1321 acttgtcgtc tttgcttctg cattactcct atcaatgttc ttcactggcg tcgattccac
1381 aaggtatcga aatgtttttg agtcaaattt atacgtgatc tggttcatgg aacaaagcat
1441 tgaagttctc acttctcaaa tctaagctaa aacaagaaaa ccctttttct tgatttgaat
1501 tacttgatca tgattacatg gattttgtat tatagatcta aatatggagg gagttctttt
1561 tattttattt ttttatcata gattatgtag agatccatgt gatcaaacac aagtgtgtta
1621 ttagagttta ttgaatgata atgcaacatt tcaatgtgta ttagcagtga ttaagcgagg
1681 tttggtatat tggttttata atttgagttg tgttttcggg tttaactaat gagttggtta
1741 attaatgact tccaggtcta atgaaacatg gcatgaacat gcagttgaga atccagagga
1801 agtagctgcc atggtggaca tgtgagtttt tctgttgctt aaatttttgtg ttatatacta
1861 tatatgaata ataatgattg ttgttttgttc tctaattaag tattttagca tctaatgcga
1921 ttggttagtt agtaaaacaa tttaaaaaat acactaaact agcattactt gacgttttgg
1981 ttggataaat tcatgtgatc ttcttttata cattttttaac tactggagaa aaagaaaata
2041 aaaatcagaa aactggtgaa agttgtattt atgtggagca gggctcacat ggcacttgtg
2101 taagaatttt tatacattct ggtcaaaagg tcaaaaggtc aaaatccata gggccaaaca
2161 ccacaccteg attccatttc tttcatactt tttgtttatg tttttgtttt atattacctt
2221 aattctcata atgataaaaa aaatttctct tatgcaaatt ttaactgttt ttaactaaaa
2281 gtatgttctt ggttccggcg cactgacttt ttcaacgaga aaatcttttg attatttaga
2341 atttttcgtt tgccgtagaa aatgattgaa aactttgttg attgttctat tttatacagt
2401 attatgtttc cacatttatc aaatacaaaa cgaaaaccta aataggttat aaaagtttga
2461 ttttcactta tagaaacatt cccatcggag ttatttttttg gttcttccgt ttttcttttt
2521 gtgttttcaa gattgactac attttgactc cgtaagttat agaacttaat gccaacagac
2581 tcgatttttcg atttgtcttc gcttccatgt ttttaattaa atttttttcgg catgaaacca
2641 atcaatacat ttaaacaatt tgcaaatcat attgtaaaac ataaaaaaaa taaaaatcaa
2701 ataagtagtg aggataagtt ctttatgctt attcttactg agggggttca gagttattgg
2761 tggaccagac atgaaacatg atgcataaca tggttcattt agttgagtaa gatttggatg
2821 ggattcttca tttaatgcaa tgcacccaaa tcatatagat gaataaattt atttttgtaa
2881 aaaacattaa atattattgc ttgcaaaaat cttatggagt tattgtagac attggtccaa
2941 aaacacgtgt aatgttcttg ttttgtttt tgaagagaaa catgtgtaaa tgttccaaca
3001 cctaatcttt aaaacagtga attgccatca gccccatgca tctccctcgt gtagtgaatt
3061 gttctctgaa tgatttcact cgtatctttа gtcttatctc gttattactt ataactatac
3121 tagtactttа aagatctgtt tggcacaagt tataatgtta tgactatatg attaaacgaa
3181 tagtattttt tggattatta tacccaaaa acatataaca ctattataac gtcatttcta
3241 agaaagttta gcagtgtttt atttttattt ttggctttta gaaatcattg gattcttgat
3301 tgaatagga tagtttgcgt aacaaagcaa gatttttttc aatttttctta caatgcgtga
3361 attagaagca ttagaataat ccgaccgacg aaactatttg tcaaatgaca cacattaaaa
3421 gtggtccacc caattcctct gatttcggtt aatatagtaa ccatgtttat atttagtgat
3481 taattctgtc ataattcaat ttagcttagt gatcatgtta aattaatctg gtttgactaa
3541 accggttttg atggttcagg agcattcgaa acagcacagc ggagaagaga ttaggatact
3601 tctcatgctc aaccggaaat ccaatcggatg atgttggcg ttgtgaccgg agatggcaat
3661 cccggcgaaa acatctagcc aattgccgca tcgggttcgg tcgcaatgca atcggcggta
3721 gagacggteg ttactacgta gtgagtgatc caaacgacga taaccccgtt aaccctaaac
3781 cgggaacatt acgtcacgca gtaattcaag aagaaccatt atggatcgtc tttaaacgtg
3841 acatggtcat aacattaaaa gaagaactaa tcatgaacag tttcaaaaca atcgatggtc
3901 gtggcgtcaa cgttcacata gctaatggtg catgcatcac gattcagttc gtgactaaca
3961 ttatcattca tgggatacat attcatgatt gtagaccaac tggtaacgct atggttagaa
4021 gctctccttc gcattatggt tggagaacta tggctgatgg agatgggatt tcgattttcg
4081 ggtcgagtca tatttggatt gatcataatt cgctttcgaa ttgtgctgat gggcttatg
4141 atgctgttat ggcttctact gctattacta tctccaataa ttacttcaca caccacaatg
4201 aggtaagaat ttgtctcaca ttttatgttt tttggttgct tgaatcggaa gttagttttt
4261 tttatggtag ctctgtttta tgacaggtta tgttattggg acatagtgat acctatacaa
4321 gggatаaggt tatgcaagtt acaatagctt ataatcattt tggtgaaggt cttattcaga
4381 gaatgccgag gtacatactt tggcctaact tatgattctt ggcaaaaact tcgaaaagtc
4441 gacaaaataa acatttttca ttaatgcagg tgtaggcatg gatatttca tgtagttaac
4501 aatgactata cacattggga aatgtatgcg atcggtggca gtgcgagtcc taccattaac
4561 agccaaggaa atagatatct cgccccaaga aaccggtttg ctaaagaggt actttcaaga
4621 gttaggtcga ctaggtcaaa taaagaaccg atcttttctg actttttgact tagtgtcaaa
4681 ctataaccct tatacttata gattacagtt tcaacaaaat tcttacagca aaaacttatg
4741 ggatgtttca ggtgacaaag agagactatg cagggcaatg gcaatggagg cattggaatt
4801 ggagatcaga aggagatctt tttttaaacg gagctttctt gcatcggttc ggatcaggtc
4861 ttggagctag ctacgctaga gcttcaagtt tagcagccaa atcatcatct cttgttggcg
```

NCBI Sequence Viewer

```
4921 tcattaccta caatgccggt gctcttaatt gcagaggtgg tcgccggtgt taatagttac
4981 taaagtctat cttcttttgg tttcttgcat tatacatggc aatagcacta ttgaatgtcg
5041 aggcttgatt taaaattgcc tcaactgtat tataaagttg taatgatttg tttcaccgcc
5101 aattattgat gtatctataa aagcaaacaa agttggaact ctccctttac tctctcatct
5161 taatgttttc ctacacaagg ataaaagaaa catagaaatt tgtcaaaaaa aaaaaaaata
5221 gcagttctct cggaaatgca atgtagtatc aataatattg gtgaagttta tcatcttggc
5281 atgtgactat ggtttttgga tcagtttggt tcggttatat cgcctatgat tgttttttgtt
5341 cttggttagc tttggttaag agtttcaaaa ccaaatttaa ccgatttttct tattccggct
5401 aaattggtta tttttggctg aagttggttt tattttccgg ttaaattggt tcaataaaac
5461 tctagtaaca gttaaaagaa gtctaaaaat ttggtgaaaa cgggtctcag tttggtataa
5521 tctggttatt aaaacaaaaa tacctaagga cggttcgatc cgaaccgaac caaaattttc
5581 cgtttagcga tgcaaaatta aagttcttcc tagaggttca accgatagaa ctgaaaatct
5641 tgattcaatc cggttcggtt ttctcggttg ggttatttat ccattggctg aaatcccgca
5701 taacgggtcg tcttattatg taggctaact acaagttggt cattcaaggt taaacctgac
5761 caacccgtag agtccgctaa ccgggatatg attaagatgg gctgaccaag actaaccccg
5821 tttggtcgag gataacatcc ttcataatga tggagaggtg tggggtatgg tgtcaaaggg
5881 aataagaagg gatgtgaagg tggtggaacc acgtgagagg agacaagaac cacatgagag
5941 catatgagag gaaataaaag aaaggacctt agtatcattg tcatgttatg ttgttgatat
6001 tgataagata ggtttggcaa atggcttgct tctctttcat tacacacttt gtttcaccaa
6061 atgttctctt tatgctaatc tctcctttgt tattttctta atttgtcaca tttagcgtgt
6121 tttttttac agtcgcagag actgcttcta cggccacaaa attacactat tacggcaaga
6181 aatttaacgc cgttgagtcg ctgcgattca gacgacaagt aaaaacagtt tttaaagttg
6241 acgctgaaat ttccagcgat tacggaatta ctgttttctt ttcttttttct gatcgctgcg
6301 gcaactgcgg cactttattt ttcccaaata cacatcatct caggcttttt atttgtgccg
6361 cagttcttgc agctgcgact gttaaacgaa catgctaatt atcttaacta aatttaattt
6421 gtttcttaag tgagtaagcg gagtaatttc ttaaacaatt tgaatagtag gttagataaa
6481 attaaaaattt aagtttagaa tttgatattc ttttttcttt tgacctgaag tgtacacata
6541 acccgtctt aaatagcatc atgaacacat caacaatcca tctcaatttg ttttcgtggt
6601 ctcgcgaaag acgtacgtac aagtggctaa caattatatt aagcatcctt taaatacatt
6661 aagaataata atcttatgtt acagaaatct aagacgtaca gagggaatta tgttacatgc
6721 tccggtttag aatatttcat cacaaaaaat atatttaatt ttgtagattt atatatgtta
6781 acatttcaaa agcatgtact ttaatctttt aacaatacat tgccgtttg aaaacaaaag
6841 tcgaaagacg tcaaatagag aattttgacg tcatactact atgttactta ttccatgaag
6901 aaaacataat atctaaatta gattttgttt tgaaaatttc gacatagtat tattgtccgt
6961 acttggacgt atacccaaaa ccacttgtat ttaaaaatca taaacaaaaa aaactcattt
7021 atctatgcga acataacaca atatcaaata tgttagataa aaagtagtca acacttttct
7081 cacccttat attgtatgaa ttatgacata atatatagtg tttatagaat tggtatgaaa
7141 ttatttggtc aacaaatata gtctccattt tttgatagtc aacactatca aaatttgtta
7201 aaaaatcaat agtttttacc catatagttt acctcaccaa ttaaggtttt ttcctcttgt
7261 agtatataag ataaactatt tgacaacaat aaagtaaatt ttacatacaa cataagaaat
7321 aatacgataa ttaataagtc ttgcaccatg tctacgaaat tattttctta tttcatgctt
7381 ttggtagttc tcttctcagg taattaactt tttttcttct tcttcttata tgctcaacga
7441 tctaatacgt caatgtacaa ttagctttga catttgaca gaacttcttt atattttttt
7501 cataggaagc taatgaacta atgctatcat ttttttcttc atacatattc attttattta
7561 aaagaagcat actaatttta atgataattc tgtatagttt ttaatttaga tacatatcta
7621 tgatacttta actcactatt tgttgcagta ttaacgatta ttcctaaaac tgaggcacaa
7681 aagagatgta gacaagaact agagccggga aagcaatgtg tgttggcaaa gtgtagggaa
7741 ctgtgtttca aacaattgaa aggattcggg agttgtattg agaagccacc aggctcttct
7801 aagtatacat gtaattgttt ctacaattgt ggtccaccag gatttttttta gaacgcaaga
7861 taatgtaaac tcaatgaaac aatagtttaa ttttctatgg attacaaaat gaaaatttat
7921 ctttaaatta caaataatgt aatagtttaa caaaaaatga aaatttatct ttatcttaaa
7981 attacaaaac ttttatactt gccatattta aggtttgttc ttctgatagt atattactat
8041 attgtgcttt aaccatttttt tgataaacgt gatgtagcca tctccttctt ctagatcttc
8101 ctgcataaca atttaatcac acaaaatcaa tttctacctc atcggattat cacagaccca
8161 tgaacaaatg atacacatca aacccaaaat tggaagcata agacaagcta aagactaaga
8221 agattggtac cctttgcgag atttgggacg agaaaatagc tagagacgaa gaatctcctt
8281 agacgaggag aatagatacg aggagatagc gagaaagtct aatttcgtt tttccccaaa
8341 tgtgtaaatt ttaaagatat tttccgtatg agctagctca ttatgcccgc gaacaaaatt
8401 ggtctttaga ccatatgatc cgttttagtt gcgggctgac cacagaccgt ccacaatatt
8461 ttggactgag cttagcccat ggttggatga ccatttgaca gttctacttta ttaaagcgat
8521 acagcttgta cagctttgac agttaccaat cacaaaaact ttatcaaaaa cgttacaaaa
8581 agttttactc agagttttttt ttgtacaact tttttgtaca actttttcta acagctttgc
8641 actgttccaa taaagctaac agctaaaaat ctttgtacaat ttttggttac taatcagacc
8701 cattatgtta gggttctccc tacttttgaga aaaatcagcc caaatgtcta actctggtaa
```

NCBI Sequence Viewer

```
8401 taaatatttg aatgagttta ttctactgta tatattttac acacataaaa cttggtttct
8521 tataaatttg aagttttgat atttttttga ataaaaactt tgagacataa ttcatcattt
8641 accttttata ttacaattca tcaagatttt tattttatca catttctaaa ctactgaata
8761 aaagattcaa ttcttataaa tgtggagtca gtaaagaaac aacagagctt gtgtatacac
8881 agggaaagag tcttccgtct ttatgtttag tttcaccaat aatccataca cgagcatgag
9001 ccatcttgga aatggtgaaa acgtttccgg tctcgtaccg ttattttct cttgtaaact
9121 ttggagatga gtttgatgga cgaatgacaa tcctcacaga tacgtaggtt cttaacaatg
9181 cgtattggtt tgccagggct cgtgcttatc aacccaaacg ctgtagcgat ctttctgctg
9241 tggtaatgaa gtacatttttc tttatcttct tcctccaagt ttaacaaaac ctgtgtggtg
9301 attggtctat acccagctag tctcagtttg tctgagattt ctacgagcat tgagttgatt
9361 tctttggctt ttgggtgtga atcatcttcc acaacaaact catgaagcac gccatcaatg
9421 tcgatcaaac tacacccagg gtcttttctt atatccttct ccttcattct cagcctcatc
9481 tctgaaacct ctgaccagtt tccttgagaa gcatacatgt ttgagagggc gacataagct
9541 ccactgtcat gaggaaccat atccatcaga atattggcaa cgcgcttgcc catctcgaca
9601 tttccctgca ttctgcaagc acccaaaagt gctttccata tcacatcatc tggtttttata
9661 ggcatattaa gtataaactc ctcagcctca tcaagaagac ccgaacgacc caataaatca
9721 accatacacc cgtagtgttc aatcctaggc tctaaaccat ccacactcac catctgactg
9781 aaatatctcc ttccttcttc taccaacccg ccatggctac aagcagtcaa aagattaatg
9841 taggcaacat cactaggtcg aactccagcc tgcctcattt tacaaaaaca gtcaatagca
9901 tcacctgcct ggccatgtat agcaaacccg ttaatcatcg cactccaagt aatcacgttc
9961 tctctaggta acctctcaaa tacatgaatc gctttctcaa tgatcccaca tttcgaatac
10021 atatcaatca aagcagagcc aagcacatca caatcctaa ttccactatc ctctcgcatac
10081 aaatgcaacc attctcccaa ctctaaagac ccaagccgtg aaatagcagg aagaacacta
10141 accaaagtaa cataattcgg tctaatatcc cctttcttca tctcacgaaa aacctcaaca
10201 gcatctttga aaaacccgtt cagcgaataa cccgatatca tcgtattcca cgaaacaacg
10261 cttctttgac gcatttatc aaacaacatc cgtgcagctt tacaatcacc aagtctcata
10321 taaccatcaa tcataacatt ccacaaaact atttcaccat ctctcttct tctatcagtc
10381 atcaccacca tatctttctc tataatgttc ttataaaaca acacacgagc atctttcata
10441 aaaccacaca tcacatacat tctaacaaga ttactcatca caaattcatc accaccaaat
10501 ccatacttca aagccaaacc atgaatctgc ttacctttct gaatcttcac agttttcgca
10561 caagccttaa gaacagaggg aaacgtaaac cgattcggtt caacaaactc atcactcatc
10621 atctcgtaga ataatgtgat tgcgatcaga gctttgtctt cgtcactctc actgaaacct
10681 ctgattatag tattccaaga gaagcaattt cgttgaggca tttgattaaa gatcttatgt
10741 gcgtagtcaa ggtcacggtg gtggagatcg gaggtagcac agaatcggag aatctctgcg
10801 gcggcgagag tgtctcgcat ttgaccagat ttaatgaaaa cggcatggat ttgagagagg
10861 tctcttatag ttcgacagtt gttgatttga ggaaagagtg aagatggatg actcgccgga
10921 gaagacgccg gcgagttacc gccgggagag aaaagtgtct gtgttgggtt cattgagaat
10981 caacaaagat tcctacgtgt ggctatgtaa gacgcataac aagatataat aaagattgtg
11041 gtcgaaaacc gattaaacgg aaccgaataa accgatttat tcagttatcg gcgctgaagc
11101 tgcaaggagg agttaaccga tcggcgaaat tatttggaa ttctcggtctt tggatcggtt
11161 cggttcggca actctgaacc gtttgttgt catatctaaa gatagttaga actctaaccg
11221 aataaccgaa taatcttatg catttgcgca gacaaagaga tttgattatt tatgagacaa
11281 gaggaagcca agttgggact aaactaatgc aaagcttagc atgagagtca cgcaagccga
11341 agcttagaag cacattccac aaatccgacc tttacactac tactattact actacaaata
11401 ggcttcaaca ttagagacta tgcacttcac ctttgcaaca accatttaag tacatacata
11461 ctcccaaaat ggctgcctct gtggaatact ttccctatta ctcaccaccc tcccatcaac
11521 atcctctccc ttcttccagtt ccaccccctc cttcccacat ttcccctcca ccacctccat
11581 tttctcctcc tcatcaccca ccaccaccac atttttctcc tcctcatcag ccaccgccat
11641 caccatatcc tcatcctcat ccgccaccac catcaccata tcctcatccc catcagccac
11701 cgcctccacc tcacgttctt ccaccaccac caccaacacc agcaccgggg catcatgtta
11761 tcattgtggt tgttatctct cttggaagtt tgttttttct tgcattcctc gctgctgctc
11821 tcttctgcta tctcaagaaa agaagaaagt ctttctacaaa ggctgagatt atcgagtttg
11881 atgaacacct caaagtccaa gaaaccatag tgcaaggacc ccatggcgaa caaacgagag
11941 tggtaatgct tgaagaggac attcatttgg tagaggatat tcacaagact gaaaagctta
12001 gccgtcctag tcacctcagt tcaactggaa gacatgccat tgatatttca gatcccaatc
12061 atcattttac tgagcaaaag tcctaaatca aaaggctcaa aatgtaaaca atcatattta
12121 ttcacaccat ctaactaata aaaacaattt aaatatggta ccaaatcttg caatccaaaa
12181 cttgatattg attccaagat aaatgctgta catgtccgtc caattaatac aacttaatat
12241 tttactatga tttccacttt gcgaatgtaa gtaacctat atatttatga tcattgattc
12301 cgtttgctgg cttttaatgt agattcctag tgtactctgt ttggtttctc aataacatta
12361 ttccatttaa ctcgagtttc agatgtttct attaaataag taatgaaagc agctctaatt
12421 ctcccaaaat tcacaagtat agaaagatag aaagatagaa acagagttat tgcatggaag
12481 ctgaaaatca tacagcatgg tggatataac caatgagaca caaaacaatc acggcatgtg
12541 ttatagaaca acacttatca tctgtattaa cataaagcaa aagttcctat acgacatagt
```

NCBI Sequence Viewer

```
12601 ccaaagttac agaccataag cacacctttat gatttgttgt accctctaaa acacaggatg
12661 atgacaagaa gcaaatttgc ccaaatttac tgagggaaaa acgagggctt aacgctctca
12721 tatctcaaac aaaaacttct caaacagttt catgtgttca gattggaggg caatggcaac
12781 ggataagctg ccatcattag taggacttgg tagcacaaaa gacaaaccct cgtatggaat
12841 tccaccaggt cccataaaga taggacgacc ccaaccaaag tctgcatcat aaataggtaa
12901 tctaacccag cttgtgattc ccaaatttgg gcacttgtag gtatgtgcac cgcggacaag
12961 ggctgacaga tcaggctgca tctccaggta gtcaagagct gacctcagat agttatcatc
13021 catacgaacc aagaaatcat gaatctgtcc tgcagcatac catgttggct tagataacaa
13081 atcccctgca acagccaatg gtgttgcagt gaatatcaca ttcccaaagt aaccaggagg
13141 cagctgcgga cgtagtctag accttccatc agttgcaatg tacagtttcg tctcttggtc
13201 gtttggaagc cctcgcgcct ttcccactga tctccacaca tgccctgcca acatctcgta
13261 tgagctgtag ctgacagtgt tcccatcctc cttggatttc gccttaagag caacaagctg
13321 gtctcgtgtt aatttgaata tagagacagt ggtattctca ggtcctgatt tagacggatc
13381 aagaggtatc ttcatacttg gtgcaggctg atattcaaca tgatgaaaag caggctgtgg
13441 cgggtccata gctcggagga gtgttcgatc aatgaaaggt ggaatggtta ggtcaagacc
13501 acgagccata tcagaccatg tgttgataaa atgaagacca gagaaccatt ctgccgcgtg
13561 atgttgcatc ccaaccccaa gtgaagctcc cccacattta aagaaagtca cctatatcat
13621 caaacaatat cattacataa ctaaaaacag aacatacatc atcaagagat atcattacag
13681 aactaaaaaa cagaacaaca caagcattca ataaattaaa aaagggaaag actttaaatt
13741 aagctgacct gcaaaacgag aagcgggaaa gagtgaatgc cagcggagtg atcaacttcg
13801 ggaataagct gacggagatt aagggtagga gcaaaatcac caaaatcatc gataacagaa
13861 ggagtatcag ccacaacgaa gagaacacca gcaccgttac aatcgatctc aatacgacca
13921 tcatcgtctc tcttcaagcg accagccata gggtaaaaag ggacaagggc tttggaaaga
13981 gcttccttca tgacctgagg gtcaaagaaa ttggaagcgc cggtgggtct gtagaagtag
14041 acactagggg tatggaatct ggggatgaca aggtcgacgt tggagttcca aagattagtg
14101 attggtgtct cggtggcagg ccggaccatg gtggaatctc tgatgttaat tttcatattt
14161 ttaagaaaga tcaaacgaac aaacaaattt tgttttgagt ttgatctgag aaattgtgtg
14221 agaagcgagg aggggtagat ataagtgggt gtgaacaaaa aaggaaaggt cgtaatcctt
14281 cacctaccgc cccacgaccc cgcctactcc atctgtcgac caactctctc agaaggttta
14341 tcattggtcc atctcatcac tcggtcctac ttttctatcc ggggttggta aacccgttcc
14401 ctctgtccta aggccccaaat aatactaatg ggcttaaata gtataaagtt cgaatgaaga
14461 agaaggaaaa aaaattgaca agttgattgt tatagctacc gtacaagtta gtcaatcgaa
14521 tgccaaagtt tggcttaact taactcatta ccacgtagca gagcataaca acaaccacca
14581 ataacatcta acggcgtaaa ctcacgtgtt aagagaatta cgtcagaatc caagaaagtc
14641 ttttttttcag tacaccacaa ataaattatg ttgaatttat atccgaacta tactatatta
14701 tatgaaaaca tacctcaatt agtcataaat ttcacatttc atttaattt attaactaat
14761 ggttgttgtt gcggatagct tttctttgct ttctgttcta aaattatttc ttagccatta
14821 tttgatgatc caacttggtt tgaggatcta ataaaaattc aatttctttg aagcattcat
14881 tggcttaatg atccaaacct tttgtcctcc cactctctaa gaaaataaag caaaagcaaa
14941 actaaaagaa aagcaactag taccgtgggt tggtcaatgg taacccatgc tacccaatac
15001 caatcaaagc ttaaatgaca atttagtacc ttgggttggt catgatttag agcggaacaa
15061 atatacccata catcaaacga ggatatacag agaaaattca tggaagtatg gaatttagag
15121 gacaatttct cttctgggct acaacggacc ggcccattcg ctcatttacc cagaggtatc
15181 gagtttgtgg actttgatg ccgctggaga ctattggcat cggattgaaa aaaatgttta
15241 cttcgttgtt aacaatttcc tgaatgcaat attttccttg tcatgaatat ttaaacttgt
15301 tattactttc ttttagctta ggtgtggaca attatggagt ttacttcaaa cgaggaagaa
15361 tcttaaacgc tcggttcagg tctcgaaaac aaaccaactc acaatcctga ctcattaggag
15421 gaaaacaatg caaaaccaca tgcatgcttc catatttcta tcataatctt ataagaaaaa
15481 acactactaa gtgaaatgat tctgtatata tataaccaat gccttttgtt ttgtgatatt
15541 ttatgtatat ataactatty actttgtca tctatggata gtgtctcggg ctcttggcaa
15601 acatatttca aagaaaagtt aatgactgta attaattaat ctgaagctag aaacagaacc
15661 ccgaggtaaa agaaaaagac agagcacatg aagtttagta cttttatata tttaatatat
15721 cattctttct tattgcttat ctctaaagca aaaacttccc taaaccctaa gccaaaggac
15781 tcagatcgat gcagaaccaa gaaggcttgt tttggatttg agagccaaat gcaaagaaaa
15841 aaactcttat gtctttgcaa atgccgattc caagaaaaaa agccttaact gttttctcatt
15901 tttccattac attgtctctg tttttagctt tcttcatctc ctcgacctct gcatcaacca
15961 atgaagtctc agctttgatt tcttggcttc acagctcgaa ctcgccgcca ccgtcagtct
16021 tctccggctg gaatccttct gattctgacc catgtcaatg gccttacatt acttgttcct
16081 cctcagacaa caaactcgtt acagagatca atgtcgtctc tgttcagtta gctttacctt
16141 tcccctcctaa catttcttcg tttacttcac ttcagaaact cgttatctcc aacactaatc
16201 tcaccggagc tatatcttct gagatcggag attgttcgga gcttatagtt atcgatttaa
16261 gctcaaatag tcttgttggt gaaatacctt cgagtctagg gaagctcaag aatcttcaag
16321 aactttgttt aaactccaat ggcctcacag gaaagatccc acccggaactc ggcgactgcg
16381 ttagcctcaa gaatctttgag attttcgata actacttatc cgagaatctc ccgttggagc
```

NCBI Sequence Viewer

```
16441 tcggaaagat ctcgactctt gagagcataa gagccggagg aaactcagag ctttcaggga
16501 agatcccgga ggagatcgga aactgtagga atctcaaggt cttagggtta gcagcaacga
16561 aaatatccgg ttctttacct gtttcgttgg gtcaactaag caagctccag tctttatctg
16621 tctattctac aatgctctca ggtgagatcc ctaaagagct tggaaactgc tctgaacttа
16681 tcaatctgtt tctatacgac aatgacttgt ccggcacgct tccaaaagaa ctaggaaagc
16741 ttcaaaacct agagaagatg cttctatggc agaatаatct ccacggacct atccctgagg
16801 agattggatt catgaaaagc ttaaacgcca ttgatctctc tatgaactat ttctcaggaa
16861 ccatccctaa atcgttcggt aacttgtcga atctccaaga gcttatgctt agcagcaaca
16921 acatcaccgg gtcaatacct tcgattctaa gtaactgcac aaagctтgtt cagttccaga
16981 ttgacgccaa tcagatttca ggtttgattc caccagagat tggattgctt aaggagctca
17041 acatcttctt gggatggcag aataagctag aagggaatat tccggacgag ttagctggtt
17101 gtcagaatct tcaagctctt gacttgtcac agaattatct caccggatca ttacctgcag
17161 gtttgtttca gcttcgtaat ctgactaagc tcttgcttat atctaatgcc atctccggtg
17221 ttatcccact ggagatcggg aactgcactt cgcttgtcag attaaggctt gtcaataaca
17281 gaatcaccgg agaaatccct aaaggtattg gatttcttca gaaccttagc ttcctcgact
17341 tgtctgagaa caatctctcc ggtccagttc ctttggagat aagtaattgt agacaactcc
17401 agatgcttaa cttgagcaac aacactcttc aaggttatct tcctctgtct ttatcatcgt
17461 taacaaagct tcaagtactt gatgtcctct ccaatgactt gactggcaag atacctgata
17521 gtctaggtca tctcatttca ctcaacggc tcattctaag taagaattct ttcaatggag
17581 aaatcccttc gtctctcggt cactgccgca atcttcagct tcttgatctc agcagtaata
17641 acatctctgg aactataccg gaggagcttt ttgacatcca agatctagac attgcgttaa
17701 acttgaggctg gaattcatta gatggcttta tcccggaaag gatttcgcgc ctaacccgat
17761 tatccgtgct cgatatttcg catаacatgc tttcaggcga cctttctgcg ctgtccggtc
17821 tagaaaaatt ggtttctctg aacatctctc ataacagatt ctcaggttat cttccagata
17881 gtaaagtgtt tagacagctg ataggagcag agatggaagg aaacaatgga ctctgttcca
17941 aaggtttcag gtcttgcttt gtaagtaaca gttcacagtt aactacacag cgtggtgtgc
18001 actcacatag actcaggata gccatttggat tgctaatcag cgtgacagcg gttctagcgg
18061 tattaggcgt gttggcggtt atacgagcta aacaaatgat tcgagatgat aacgattcag
18121 agactggaga aaatctatgg acatggccaat tcacaccttt tcagaaactc aacttcacag
18181 tcgaacatgt actcaagtgt ttggtagaag gtaatgttat agggaaaggt tgctccggga
18241 tagtgtacaa agctgaaatg cctaacagag aagtcatcgc ggtgaaaaag ctctggccag
18301 tgacagtacc taatctgaat gagaaaacta agtcatcagg agttcgagac tcattcctag
18361 ctgaagtaaa aacacttgga tcgatcagac acaagaacat tgtacgcttc ttgggatgtt
18421 gttggaacaa gaacactaga cttcttatgt atgattatat gtcaaatggg agtttgggaa
18481 gtttgcttca cgaaaggagc ggtgtatgta gcttaggatg ggaagtgagg tacaagatta
18541 tacttggtgc agctcagggt ttggcttact tgcaccatga ctgtgttcct ccccattgtc
18601 atagagacat caaggcaaat aatattctga ttggccctga ttttgaaccc tacattggag
18661 atttcggact cgctaagctt gttgatgatg gcgacttgc tgttcttcc aacaccattg
18721 ctggttccta tggttacata gctccaggta cttcccaaat ttctgttttg ctgctaatta
18781 gttaacttga tacatcataa gttactagta gtcttagcct agtaatggta gaaattcagg
18841 tttatagttt ggttttggtt tttggcattt tgaaaagaat agtattttgg ttcagttgat
18901 tttcgatttg gtttaaaatt ttagaaaaaa aatggtagaa gaaataattt ttgattaata
18961 ttatatcatc tactgtatac gatttaaaaa aaaaattgat ttggtttggt ctcgatattt
19021 tgcttgtagc ttggttcaat atttttacta tatatatgat tcatgacttt catactaatc
19081 aatatttgtt cggtttcgtt ttgattagtc tgaaagtgaa tcatatagtt aaaatattga
19141 ccaatctatc atatacttac acgctacaag atttгgatg taatttttgt gtaagcattt
19201 tggtttttat atgttcatcc ttagtaacta acaacaaata tcaattggga agcataacgt
19261 tatatattcc cataaggcta taacactata attatgcaga atacggatac tcaatgaaga
19321 taacagagaa aagcgacgtg tacagctatg gagtcgtagt gctcgaggta ctaacgggta
19381 agcaaccaat cgatccaacg ataccagatg gactccacat agtggactgg gtcaagaaaa
19441 tcagagacat acaagtaatc gaccaaggac tacaagcaag accagagtca gaggttgaag
19501 agatgatgca aacgctagga gtcgcgcttt tatgcattaa tccataaccc gaagacaggc
19561 ctacaatgaa agatgtggct gctatgctta gtgagatatg tcaagagaga gaggaatcga
19621 tgaaagttga tggttgctcg ggaagttgta agaatggaag agaagctggc aaagatgatt
19681 cgacttcatc ggttatgcaa caaacggcta agtatttgag aagtagtagc acgagtttct
19741 ctgcgcttcc tttgctttac tcttcttctt ctttctgctac ttctaatgtt agaccaaatc
19801 ttaaataaaat gcatgggtaa ttatgtttтt aagtgaacaa gttttctatgt ctaaagtttt
19861 tgttttatct atatatgacg ttgagatgaa ctcggtgttt caaaccaaga aaaattaacg
19921 gagactaatc tcgtcttacc ttttacattt tcttccgtca aggcgtcaag ttgatttttt
19981 tttgaaaatc aaataaagaa aaaactagtg aaaacagata tgaggaggtc atаactcata
20041 actgtgactc acgagataaa tcatgtatcg tggctgacgg tcacggaacc ttttgctttg
20101 tacccatttt ttcgaggttt gtccaggaag gcggtataatt agtgtacctt cactatatac
20161 tctatаactt aataatttta tatcaattag atacaaatgt ttaattaaaa aataaaatag
20221 caaaccctta accaagacca gtgtgtatat taagtaacga aataaaattg aacagaaatt
```

NCBI Sequence Viewer

```
20281 gttgaccaga taaataataa gtatagattc atgttttagt ttttttttaa cttcaatttt
20341 cagatttttt ttgtagtcta aagtttggtt gatgaaaata aaataaaata aaagaaaaat
20401 ctatatttac ggaaatatat ctcttacaaa taattctttg caaatgtgga agacctatta
20461 tacatataga catatatatt aacacaatca ttatgagtaa gtatcttaat attataaatt
20521 gatgaaaatg ttaaaaaggt tataaatcta tatatatttg tctatttaaa aatggttgaa
20581 aatcttcttt tcaaaaaatc aataattgtt tttatgaaac agtgaagctt aatctttaag
20641 aagtccctca aacaattttt accataaaat aagatttcaa attataacat acctaatagt
20701 tttatctata tattctaact atccatgttc ctatggtata ataatatctt ctaacattat
20761 ttttgaaact tgtttctctc tagtttttct tttttctcaa agcgagatcc aagttttattt
20821 caagagttac caacataaac aatggcaaca aataaatttt tatctatttt gttgctttcg
20881 ttgatggcgt ttgctgctat tctgctgcca atgatttcag gtaacacatc tattatctat
20941 taaaatttgg tttgtaatgt tatgataaat aattctactt taaatatgaa agttaccgtg
21001 tgtttctatg aaaaacaaaa cctaacctaa tttttgttg aaggttttaa ttagctcatg
21061 ggtgttaaac ataaactatc aaaaaatata accaaaaaaa taaagttttt aagttgtata
21121 tttacttttta tgtgaaaata tttctactcc actacgacaa aatgtcactt ttgtaggttt
21181 tttgtgaagg tcacaaattt tagattacga gtgcaattag agtaacacaa atatttgatt
21241 ttgaactaaa cagtcgtggt tttaaaaaata gtgttcaaaa tcagaagaag aaaattgcac
21301 tatgtagtaa gaaaatagtc gtttattgaa attgatcaat ttttatttta gagtcgttta
21361 atttgtttgt ttataaacta aacaggacaa atcataacat gtttaccggg cgaatgtaca
21421 aacccatcgg aatgcaatgc agcttgcaaa tcaaatggat acaaaggagg ggcttgtgta
21481 agcatgagca taggctcaac aaccggtgct tgttgctgta aacctaattt caaatctcaa
21541 gattctttca aatctaatga tatcattatt aataataatt agtacatcat tagatatatg
21601 tattttgata tgaataaaaa tatattcaaa tatgtttcga tattgaacca tctataataat
21661 ggataatcat tcgtaagtta tagtacgatc caaacatttg atcgtatata ttaataaagc
21721 atacgaatgc tgaaatttat gttaactata ctcatagaaa tttatttata tctgccatac
21781 caatacaatt taaaagattt ttcaattaaa tgagagataa ataatttatt ttgattattt
21841 ttttgtttct gttttgtcaa caatatgtaa tatgccgatt ctaaaaaaaa aatcatactc
21901 tatatcatat attgtctaac cctctagtta gttcaaacaa catagtaag ataacaagaa
21961 ttaaaaataa atttaaatga ttttgatatc atatattgaa ttaaaatttc aaatttctatc
22021 aaaattcatg attaaattaa cacatccaaa tgccgaaaaa aaaaaaaaaa aaaattaaca
22081 catccaaaaa ttgtcacaaa tcgtactaaa ggcccagtaa gtcttttttgt caataagtcc
22141 attaagccca acaaacctct tcttctcttt cgttttatga catcaatcca tttgaagcaa
22201 acaaatcaga cgtaaccgaa caaaaaaaaa aagttggttt gatcgatcgt taaaacctaa
22261 gaaatggatc caaaatcgcc ggaatttatc attgatcagc cgcttaaaat cttaggattt
22321 gtattcgatg agctctccgc caccagagtc tccggccacc tcaccctaac tgaaaaatgt
22381 tgccagccgt tcaaagtctt gcacggcggc gtatcggctc tgatcgctga ggcactcgcc
22441 agtctcggcg ccggaatcgc ctctggtttt aaacgcgtcg ctggaatcca tctctctatt
22501 caccacctcc gtcctgctgc tctcggcgaa attgtcttcg ccgaatcttt tccagtctca
22561 gtcggcaaaa atattcaggt aacctataca taacactgac caatacacat atatacgcag
22621 ctatctttag actttaatgt atataaaatt atattaacta gaaaagtttt tagtaatatt
22681 aaattttaat attgtgtcaa agtaatgtaa aaataaattt tttttaaatg taaaaaaaaa
22741 aattaatatt ttgttgaaga agttaaatat aattctgcaa tttgagagat aactaagtat
22801 aaattttcac gtattagtaa gtaaaagtga caagttatgt tttataacca ggtgattttt
22861 aaaaaatgaa atttgttttg atgccaatatc gtgtgaaaag taagaaaact aatatatgca
22921 ataccatatt taaattgtga accttattaa ttcacctaa aatgaaaatt ggaaaacttt
22981 ttcatgtatt tgatagactt gtgagtttga attgtataaa tttccggcaa gcaaaagcag
23041 ctcgtgtacc aagttgtcgg aaacttaat ttgttagca tgcattggac taaaccaacg
23101 caaatatttt attttattta tcgtagactg aagacctaaa gagtaaaatg aatgtgaaag
23161 tgacacgttg tgtggagtat tttgtatcta actaggtgtg ggaggtccgg ttatggaaag
23221 ctaagaaaac cgagacaccg gacaacaaaa taatggtatc aacatcacgg gttactctct
23281 tctcgcggtt acctatacgg gaccatgtta aagatgctcc agacgagctg aagaaagtta
23341 tatccaaatt ataaattcaa attgtatcta tctcattttg ttaggaataa gaacaatttt
23401 agttttggttt tattgtacga accaaacgtt ttaataaatt gttgtaccaa acatgagata
23461 ttataaggtt ataaaaattg acaagttatt gtgggtttag tttgtagatc attttgaaat
23521 atagtattaa tagactattt tgttgaaacg tccaaatcgt atagatgggg taacaaaact
23581 aaactccttt atatgattct tttgcataaa attctgtctt tgtaataact gaatcataca
23641 tacttttttta gtttgatacg catatatata ttaccacaac tatgtagtgc atttaagata
23701 tctttcttca tattataaat gtgtatcaat accatttaaa tcttcgtata caaagaaaaa
23761 aaacaagatg gatattagtt tttcgatgga gatattattt ggttgagttc tttcttcttg
23821 agacaacaac atccagtgtg gagtctaaat gttccacaat accatccatt tgtatacccg
23881 tggcttaagc aaatctgatc acatacgctt gtacactcgg accgataaca catggcgtca
23941 tgaacgtcaa tagcgtttat tttagctgta aaagatatca taaaaatcta ttatatatga
24001 tagatcatca tttatgttta tatttagaga gctctacatt gagattcgat cacgatcaaa
24061 cggtaaggtt acttcacgat aaactaacga agatagtgta ttttgctgaa tgatgtaaat
```

NCBI Sequence Viewer

```
24121 tatattaaaa aaaaatataa ggtatatata taaacaaacg aaggtaatat ggtttagatc
24181 aagaaaatac cctaatagaa atcaaaaagt ttgtgtatat gttttatgaa tttgtaaatt
24241 gacttaccgg ccatcgcaaa caagagaaca attgttatgg cagtcagtat atgtgaatgt
24301 ttctggttct gcattttttt gttttgaggt gtcactagct aatagctatt actttgtttt
24361 ctctttgttt ttagatgtat ccattgagat taaagtgctc ttatataggg gtttggtttt
24421 cgcagataaa attaatcaat ttagctggat tttattctat acccacaaaa taaatttata
24481 aaagctaaga tataattttg atatgatggt tcatgaaagt gtcaactttg gttagattta
24541 aacctgcata acgcattatg aaattgacta tatatatatt aagaaagaga taattctcat
24601 ttaatttttа gtaaaataat tactcaattt tatattatac cagaaacaaa gtgggttcaa
24661 aaattattaa ggctatatca tatgtaaata caattaagaa acatttatgt aagaaaatta
24721 gaaactgaaa tttcatatca agaattaaca aagaaaacat taataaaatt tcttgttacc
24781 ttaaatatac catttтattt aaatgaacat taaatcatag tattttagat ttaagaaaaa
24841 agaactacac agaaatttct ttaataaatc taatatatat gaaaatаaat agaaagcact
24901 actgaaagaa gatccatata aattaaagat agtttggtca atgccaatca tcattttgaa
24961 ccaaatccac atttaaggaa aattgatgtt attctttaac ccgaaaacat gttcttgtct
25021 ttgcacatgt gtctcgccct atttttтatt tcttcattca acaagtcaaa atgaatctat
25081 tttactacat gatctctata aacgaatgtc ataaagattt tatggccttc aaacttggtc
25141 tcaacatcga tcatgtatat gattatctga ttcatccata tagaagaggg ctaacccgcc
25201 ccttgtacct tgctttggcc ttaaaaaggt tttcaagaat ttacttтctt tcttatttga
25261 ttttaacttt gtaaatgtca cattttтccg ttaattaatt ctttcgctga gtgacaaata
25321 tgttaatatg acatttatca tcaagatata tattttatca tatttttgat acatgactat
25381 caaacaaacg tcacactatt ggtcagagat tattcgacat aatccataag aagttgtaaa
25441 tatatcatca agtgtgaata aatgtatata aatattatta aaaaaaaaaa ttagtagttg
25501 ctttttcgcg atgcgtgcga atagattctg atgccaaatt tggatattat taagagcaca
25561 tgtttctcct tgtgtcgtgt cttttattta agagagaatc atcatctctc actctcatct
25621 tctagaagtt aaaacatcac gccatcttcc gaattcaaac gccgatgttc tgtcgttttc
25681 tgtctcacat attggaacga gcgtaaagat gcctaatttc tcagttaacg ttccccaact
25741 ctcatctctt tacagtacaa aaacgcccaa agtgagaatg aatctatgtg ccgatcaggt
25801 gttcgataaa aagcttctgt ggagagatat gtcaacgaag atgaaatttc cttcttttтc
25861 tgctgcggaa ttacctgatt tgaggaaaag taacaagagg aggggatctc ttaggatgat
25921 caagtgcaga gccgccggag ctgacggtgg acgcgtggct gttggggatg atgtgttttc
25981 ggttactact tcttctaagt atgaagttga ctatctgggt caaagtacta agggagattt
26041 gaatctcaag cttgaccctc ttcagtcatt tggtatatat acaatcttcc tttttgtttt
26101 gttagttttc gaaattcaga ttggttttag tgaaatgtgg gataaagttt tggtctttgg
26161 ttagggaact tgattttggt tgatgtttgt ttaattgaga taaataagag ctaaaaagat
26221 cgaaccttcg ttgtgaatgt cacgttgtat gagcaaaatg taactttggt gtatctatgt
26281 aggagatggg caggctacat tggagggtcc cattgaggag gtagccgagaa cagaggctca
26341 agcggctgaa aatttgatta gagagttggg tatccaagtt cgttatctaa tccttacaac
26401 tttgtgttat gtatttgaag tttgagtatc cctttgttgc gaaaacatga atgtattgag
26461 ttatttgtgt ctggttttat atggacttct ctttttcttt tgtgtcacat tctcatgcct
26521 ggtattgcag ggccctttct ctgcacagca ctctcctcgg ggtatatttt gtagtcgtac
26581 attgaatctt cggtccatta gtgcaattgg atatgatatg gattacactt tgatgcacta
26641 caatgtcatg gtgagttctt cctttctcag acataagtag gaaacattta tgcacccata
26701 tttgatgcta atttatggtc ttgttcaggc ttgggaagga aaggcttatg actattgcat
26761 ggaaaatcta aagagcatgg gtttcctgt tgatggactt gcttttgatc cggaactggt
26821 gatatgttta ttttcttgtg aaagttgaaa ggttccacat tgtagtaccc tatctagttt
26881 ttagagtact caaactatgt tttctcttat tatagattat cagggtctc atgattgaca
26941 aagagaaagg taatttagtt aaggccgata gatttgggta tgtgaagaga gccatgcacg
27001 gtacaaagat gttatcaaat aaagctgtca ggtatattg actttтcaac accattggta
27061 atgtgaagaa aaaagagagt tgcaaagtca ttgttctgtt ttgaaggaag tctcattctc
27121 ttattgtctt tgtatggatg aaagtgagat ctatggaagg gagttagttg accgtcggaa
27181 ccagagtcga tgggagtttc tcaatacatt tttttcagtt tcagaggctc tggcttatgc
27241 acaggtacgt tcttttatta caagaaatgg tagctgaaca atttgcagtt gatttggac
27301 ataatttact ttatttattt atttgtaact catatаattt tcacatctcgt ttttcaact
27361 gatgactctt caccccaatg aacttтcttt tttacaaaat tttccgcacg gctccctgcta
27421 agcaatgaaa gttcaaatca tttcagatgg ttgatagatt ggatgatgga tttatttccg
27481 cagatcttgg cactcttgat tataaaggac tgtataaggt tgtgtttcat agaacctaaa
27541 tttcattgaa ctgaaatctt tgcatatata ttttttaatc taatcatttt ttccaggctg
27601 ttgcaaaagc tctcttcaga gcacatgttg aaggacaact taaggttgat agtttgtcca
27661 ttccataatc attttcttaa atatcagttt tctccggtcc ttcttgaata atagttacth
27721 tgcagagtga gataatgtcc aagccggaac tatttgtcga gccagaccca gaactacctt
27781 tagctctttt agatcaaaag gaggtaatgg ttcagtcttt acacaatata tctaactatg
27841 atggttтcag ttgttaatat tgtcatgttc gttcatagga tggtaagaag ctcttgctta
27901 tcacaaactc ggattatcac tacacagaca aaatgatgaa gcattcattt aacaaattcc
```

NCBI Sequence Viewer

```
27961 ttcccaatga catggactgg cgagatcttt ttgacatggt gagcaaacgt aaatttgttt
28021 ctgcatttg attgtaaatc tgttgttttt cagagagtag tttgcattag actttactgt
28081 ttttgaattc atgtttccac ataggtgata gtttctgcga ggaaaccaga gttcttccag
28141 atgtcgcacc ctctatatga ggttgtgact ggagagggtt tgatgcgtcc atgcttcaag
28201 gctgaaacag gtaatttctc attgtttcta ctgaatttaa gctgataact ttacggtgga
28261 gaaagctaaa attttttcc tctcactgtt tgtctcctat aggaggtttg tactcaggag
28321 gaagtgctca aatgatagag agttcactca acgttcatgg agatgagatt ttgtatgttg
28381 gtgaccacat ctacactgat gtcagcgtat ccaaagtcca tctcaggtgg cgaactgcgc
28441 tgatttgccg tgaactggaa gaagaggttt ggataaatta tactctgaac tatatcaata
28501 gatgacagaa cttgaagggt tttcgttcgt tgaagtaaca taaaaaccct gattttctct
28561 tgtttctttg cagtatatgg ctctaattgg cagtcgtggt caccgagaag agctaataga
28621 gcttataaat caaaaagagg ttgttgggga tctctttaac caacttcggc ttgctcttca
28681 aagacgaagc aaaggccgtc ctgctcaggt ttgtttatct ccttacaagc tctagagtct
28741 atcttctctc cctgtgtgga acacaatgtg tatgaacaga taaatagata gagatttgtg
28801 acggaaattg cttataaaga tatccgtttc tcttttcctt ctagactctc gctgctacca
28861 acttggatga tcaagaactg acagagacca tgcaaaagct tcttattgta atgcaaagac
28921 tagatgacaa gattggtcta atgctggaaa cagatggcga gctcttttaac aaaaggtaaa
28981 aagaaaaaga caaacaaatg tagtaagatt actcaagatt cacagaaagg gatatgtctg
29041 agccataatt cttgggctgt tttgcaggtg gggcttcctc tcacgcgcgg gtttgtggga
29101 taaaagccac ttgatgagac aaatcgaaaa gtatttatgc ctcttctctg catttccgag
29161 taactcataa gctcaaacca agctcacatt taaactacac ttactttctt caaaattggt
29221 ttcaggtatg cggatatata cacatcaaga gtctccaact tcctcaacta cacacccttc
29281 atgtatttcc gctcacaaga gcaggtaaaa agcttaaatc agattatcag tatagataca
29341 aacgaaatct gaggcagatt ctaattaaac tcactcatgt ttttctctc ctctgcagtc
29401 actggctcac gattctccgc ttccagatgc gggtatagaa aactagcggt tgtgaccaaa
29461 gttgaggcaa cttgtgtgtt gtaatatttt tagatgaaaa aactgtggat agaagaaatg
29521 taacaataga aattcttaag ctcaaaagtc tactttatgt ataaaatgtg tattctttatt
29581 ctctgaattg aagttaaagc tggcctggt tggtatgttg ctgttgtttt tatttactgt
29641 cttatctttt ttaagctttt tcaggctgac atcagaggca gattgtgcaa atgactactt
29701 ttagagaaac attatacttt gagatttgt gaattctgat gactgttata tcaaccacta
29761 gacactagac agtgcagact tctatataaa caatcgattc gttgtgagca cacaattcat
29821 tacttcaatt gcaagagaat taataagaaa tctctcaatc tctttattat tcagtatctc
29881 gatctttcgc agatatacaaa tttgtgatat tactatacga aacatctaag acctatgatg
29941 cagaaaacgc accgttttcc ttacgtcaaa gtttgaaaca tacggactct tctctgtcgt
30001 tctcattcgc tctctggat tttgtttctc caagcaaaa aaactccgac tttcaacgaa
30061 aactcagtgc ctcgtccctc tctttcaaaa aggtactctt ttttatcac ccagcttagt
30121 tttaggttgt caatttgga tcttagggtt ccgtaatttt aagattctcg aaacatctt
30181 ttgccttgcg ttagctgagg attagaatcc caatttgagt tggatacctа aaagtcagat
30241 tctttttgt tgtaaatttg aaggtttttg gttatggcga cgaagaaagt gatagctata
30301 tgtcagtcag gaggagagtt tgtaactaat aaagatggct cattagctta ttctggtggt
30361 gatgcttatg ccatagacat cgatccaggac acttgtatga gcgatttcaa atccgaatta
30421 gctgaaaatt tcgggtttag ttgggagaat atgacactca agtacttcct ccctggaaac
30481 aagaaaacct taattactat ctctaaggat aaagatttcc aacgcatggt tagctttttcc
30541 gcagacgcac caaatgtcga aatcttcgtc ttacccgaag aagccgaagc aaggaatgtc
30601 tccaacatgc cagctagtag gtgtgtgttc tagttctaag acctaacggt ttatcgacat
30661 tgtttatgat tgtaatgact gatgtgtgga tttctgtatt tgacagatca agtaggacaa
30721 ctgcatcaga agcagtggta cctgtagtgt ctgcaggaga tattgttgtg ggagatgata
30781 tgatagcaaa tgattttcag attgaaatgg gtataatgcc tgaagaaagc agcccttgc
30841 cgtgtaattt tgttttgact gatgagaaac aacatataaa agctgcacag cagtgggaaa
30901 acgctattac gggtgttgat caaaggttca atagctttac tgagttccga gatgcgttgc
30961 acaagtactc gatcgcacat gggtttactt acaagtacaa gaagaacgat agccatcgtg
31021 tttcagtcaa atgcaaagct caaggttgtc catggcgtat cactgcatcg aggctttcga
31081 ctacgcagct gatttgcatt aagaaaatga atacgaggca cacgtgtgag agagcggttg
31141 tgaaagctgg ttaccgtgca agtagaggtt gggttggaag tatcataaag gagaagttga
31201 aagctttcc agattacaaa ccgaaggata ttgcagagga tatcaaaaga gagtatggga
31261 ttcaactgaa ttactcgcag gcatggagag ctaaagagat tgctagagag cagcttcaag
31321 gttcttataa aaaggcatat agtcaacttc cttctttctg caagaagatc agggagacaa
31381 atcctggaag tattgccatt ttcatgacaa aagaagattc gagtttccat cggcttttca
31441 tatcgttcta tgcatcaata tctgggttta gacaaggttg tcgtcctctc cttttccttg
31501 acaccgccga cttaaactct aagtaccaag gggttatgct tgttgctaca gctcctgatg
31561 ccgaagatgg aatatttcca gtagctttg cagttgtgga tgctgagaca gaagataact
31621 gggtttggtt tttggagcac ctcaaattgg cactggccga tccccggaca atcacatttg
31681 ttgcagattt tcaaaacggt ctgaagactg cattacctct ggtatttgaa aagcaacacc
31741 atcatgccta ttgtctgcgc cacctcgctg agaaattaaa catggacttg caggctcagt
```

NCBI Sequence Viewer

```
31801 tttctcatga agcaagacgg tttatactca atgatttcta tgctgctgct tatgcgacac
31861 aacctgatgc atattaccgt tccttggaaa atataaaaag catttcccct gatgcttata
31921 cttgggttat cgaaagtgag cctctccatt gggcaaatgc tttgtttgaa ggagagagat
31981 ataaccacat gaactccatt tttgggctag atttctacag ctgggtttct gaagcacatg
32041 agttgcccat aactcaaatg atagacgaac ttcgagcaaa gttgatgcaa tcaatatata
32101 cccatcaggt acaatcccgg gaatggattg tttcaacgct aacaccaacc aacgaagaga
32161 agctacagaa agaaatagaa cttgcaaggt cacttcaggt ctcagcgcct cacaatagct
32221 tattcggagt tcatggtgaa accatcaact tagttgacat caatcagtgt gattgcgact
32281 gtaaggtatg gagactaacg ggtttaccgt gtagccacgc agttgcagtg gttgaatgca
32341 tagggaaaag cccttatgaa tactgctcca gatactttac ctcagagagt taccgtttaa
32401 cgtatgctga gtccattaac cctgtcccta acacaacaat gacgatgatg attttggagg
32461 agccgcccgt tgaaggggtt gtttcggtta ctccaccgcc cacaagacta acacccacag
32521 ggaggccaaa gagtaagcag gtggaaccgc tggacatgtt caagcgtcag ctccagtgta
32581 gcaactgcaa gggtttagga cacaacaaga agacctgcaa agctgtatct taagtattag
32641 ggagagagag agaaatagag aggagagaca tggttaggac aaatttgtag ttatatagct
32701 gactttgccc aacatttttgg ttttggttct tctgtttgtt gggtttttga gaatgatctt
32761 aaagggtaaa aatctccttc aggttgaaac ccaaacgtga agagagcaaa gcggctcaga
32821 atctctgttc ctaaatggca ttaaacttga tgaaagattt tacataaatt tttttgtatg
32881 taatttgatt tttaacaaaa aatttcaatg aatcttttca attacatata ttttgtattt
32941 cgagctttac aaaatccacc acacttgtta aaaagcttat gcagttgccg acaacacttt
33001 gacgtttaca ttccaaaaat gatcgatgtt catatgagcc tataataaat gctagagaca
33061 gtaagatata agtatccaaa gagtaaaatc tctaccaaga aattgactcc aaccaatcgg
33121 aggtaaactc ataggccaca aaggttacag caccagcagg agcagctttc acagtcgatg
33181 gcacaatgcc tttgtacaga ccatgccacc cttctgatat cataatctgt ctaagaccgt
33241 ctaacatgtt tctgtatgca cgtcgctcca ctcgggctcc atatcttgga tgtctctgta
33301 aaccttcaat ctgcaacaaa agaaactttc aaaatctatc tggcatttct attatttcat
33361 gtttgaaggc aattgcattt tcattgcgtt ggagatccac tggcaagttt gaaaagcttt
33421 tatatttaag aaatgtttca cacaaattgg agttttcttt ctataggtac ggtgacagaa
33481 caagtggggt acccggaatc agtaagtaga ctcagcsagc acatatcata gggcagtaaa
33541 tgtgaaagtc tagtctaaga tacataagga tagacccaaa acatgaaacc acatgaatac
33601 gataagtcat tgcgctagtt cagtgtagga gcattgtaat taccctagct ggtttgtagt
33661 agatagtaga tgtgttgaca actcagcaac tatagttgag taagttagca agagagctaa
33721 gtggctaaaa ggaagtttgg gatgtgtaat taggtcgagt agagattctc attaatcaag
33781 tctctctatg atgtgtagtt cttgctgaag tgaataaata aggagtttaa tagtttggat
33841 cctatcacat tatcttccca gagaatccat tgggtaaaga ttaactgacc tggaatcgtt
33901 ttttgaccac atcaagggga tggcagacga gcttggcgct agtgccgcgc ccaagcccac
33961 aaatgaaaag ctggaagcta gaaaggttgg tgtccacatt gatcgggatt ttggaggata
34021 atttgtatcg attccagtcc tgtttaaaaa taatcaattg agttgattat catatgaaca
34081 ttgatcacac ataaactata tattcctatg ctcttaccat catccaacgc ttaaacatat
34141 catatgtgcc aaattgcagg ccagcgtaag gcacaatctc aacaagagtt ggtgttaatc
34201 cattatacaa tcctcttata cctcgagact gaataatatc gacaaaagca gacctcattg
34261 ttggatacac ctgttgacag tcatcagtca gtcatctata tgtcacttaa gcatgaataa
34321 ggaagaaaaa caagaaaatc cccagtacat ggaatttaaa ttatattaca ggttcataag
34381 tcacgcaact aacctaggc tccctctgtg aggccaatat tgtcctttaa agatcaaaag
34441 gataagatcc tagagtagca gcacatcctg ctaacgctcc actgacgagg gatagataag
34501 ggctcaaatg tatatgatcc tctgcattgc cataagatca caaccagcta ataagtaagg
34561 aaccacattg atctctaaat tatacgagtc aattccaaaa gctattcaaa gaaatacaca
34621 attaacaaac agaaatgata cccctaagat agccaccaga ctacaacaaa aaatagcact
34681 ggcaatctac caaatacctat atctagagag accgcaatgt tactcagtag actaaattac
34741 ctgtttttagt agaaccagat gcaaaagatt tcagcttatg tagtactgta aactgaattg
34801 aagtatacgg catgaccatg agcaaagctg gcacattacc ccgccaaaac ccctaaagaa
34861 aacaagtatc aaacatcaaa acagagaaca agacctcata aaaactaatg ccacaaacct
34921 gaaactacta acatactata aacaaaaaag aaatggtttc agaaaacgga aatgttcaag
34981 agacggaccc gaaaaccttc ttctctaaat atatctttag tcgcctgtac cattccagta
35041 tacttagaag ctcctgacaa gtttccccga accaaacccc atgaagtcgt cggttctagc
35101 tgaacctgca atcacatata acattcttca accatctcta aatatataaa agctaaccat
35161 tttctaacaa tgataattct catctaaaac ctccacaata caatccaaag ctaacaacaac
35221 aaactcaaat attctttagc taaaccaaat aaaaatctaa actttatcac tcagaggata
35281 gaaaaaaaaa gccaaacctg aaatctaatc ttaatgacat caagaggaga agttacagaa
35341 cgagaaacac caccagagat agctccggca gaagcatcaa tcagagcacg tttaatctgt
35401 cccggctcat catcaaccgt cgccgtcgta ggaatcatat ttttctaggg tttttggtct
35461 ctatagcta aaatcaatta aaaaagtttc caaaacccta tcatcaatat aataaacctc
35521 aaacaaaata aggaagaaaa ctaaattcaa ggcaacaaat aaaatcaaaa acgaaaactg
35581 aaatcttttt taccagataa agcgagaaac gaaaagggtc gacattgatg gagtttccgg
```

NCBI Sequence Viewer

35641 ggaaaaagca aggaactttg agatgaccct taaacgtttg ccgcggattt gagagttggg
35701 tggagattga aaatgatttg actcggaaaa agattataaa aaaaaaaaga agctgagaga
35761 ttcttttctcc gtcaacgttc gtgttctctc cgatagagag attaaagctg gctttcttac
35821 aaaatttgcc ctttttagga ctcttgtcgg ttcggtttaa tgaagtatca attcgtctta
35881 aagtaattaa attgtttcaa aggggtttta tggttactct taaaccttct tcctttacgt
35941 ttgttggatg gatttagtca tttactaatt tgaagatcaa cagagaccct tacaaaaaga
36001 aaattactcc ttatgttcca atatattaga tgttttctgc tttcttcttg tttcatttta
36061 ttatattaag aaaaaacatt atgtataatt tatgcatatt taatatgttg tgattatttc
36121 tattctgcaa tcttttttct tttgattaaa ctttttttaaa taaaggtaga taagatttttt
36181 ttagagaaat acactaaaat acatttatgt aaaaataata aaatagtttc ttatttttttt
36241 taatgatttt agctaaaaca ttttcttatg taaaatatga ttttcatagt tggtaactca
36301 agtgtttgca acacatcgac atgtgtcaaa tctaattttg ttaatattta aaacaaaaaa
36361 tatagaaaat ttttaaccct atggaagaat gaaaatacat acaaaaatca caaaaagtta
36421 tatttctatt catAAATTTT taagtaataa gatttagaat aagttaatgg gatgtaacat
36481 tgctaagagt tagaatactt tctctttatt tttgggggtaa agttgtagaa taattaatta
36541 gttacaaaaa aaagttttta atttttgaa gtaaagttgt agactaatta attatttaca
36601 aaattttttt tatttttttt ggagtaagtt gtagaataat taattagtca caaaaataag
36661 tttttaattt ttggagtaaa gttgttgaat aattaattag ctaaaaagat agtatcgttt
36721 tgtaattttt atatttatct ctaaatattg gtccaacaag gcccaatttt tgattgaatt
36781 tgtctcaatg gcggattccc aaaggttgtc caccgcttcc ggcgtcaagg acggccaacc
36841 accatggaag aagaagaagc tgtcgaacga tacgacttcg aatccgtcac ttccatatga
36901 tgtaatactg atcatcttag cccgcgtctc gagatcgtat tacacaaatc tctccttagt
36961 ctccaaaagc ttccgatcaa tccttacttc accggagctt tacaaaaccc gaaccctctt
37021 aggcaaaacc gagaatttcc tctatgtttg cttacgcttc cctgatgaag ctaatcctcg
37081 ctggttcact ctctacagga aacctaatca aaccctaacc gatcacacca ctaagaagaa
37141 gaagaagaag aagaagaagg aggagaagtc aagtgttaat ctgttggctc caattctaat
37201 tctcaattct catcctgtgg aatggtcagc tatcatatct gttgatcatt atctctacgc
37261 cattagtgca gacatcgaaa aggctccgta ctctaacgtt ccttacctag attgtcgaac
37321 tcacacttgg aatgaagctc caagaatgag attggctcac actaatagtg aatttgaagg
37381 aatagtgtat ttaccaggaa gctttgaaag tccagattct ttgaactgtg ttgaagtgta
37441 caacacaatg actcagactt ggaaaccggt gccccccgaa aaacggatgt tcaaattgga
37501 aaatttggaa aaaaagattt actacaagag ctttcattta gactctaggg ctgggaaagg
37561 tctatctttg agctataaat cgaagcattt gacatgtgga ttggtagtat tggatacggt
37621 ggattcatat cttcgctctt cttgtatgat agagaatata gcctatttta atcgtaaagg
37681 aaatttttatt tggagaggac ttgacggtaa gatactggtt tacgggaaaa tagagggttt
37741 agaaggggcta cctaagtttt ctagatattc tagtgttcaa ttggcggaat atggcggaaa
37801 gctggtggtt ttgtgggaca agtatgtacc tgcgagtggg tataaggaga agatgatttg
37861 gtgtgcagag atttcgctcg aaaaacgtaa cggtaaagag atttgggggа gcatgtgagtg
37921 gtttgatgct gtgcttacag tccctaagtc ttacaaaatc ttgtgtgcta ctgctgctac
37981 tctttgattc gatgactcgg gaacttgaat ttgtcaagct atttattgtt tatttcttc
38041 tcagtcggaac atagtcatta caattcagtt aactcatttt ccattgatat tactcctaa
38101 ttttttcaat agttcaaatc aagttagatt tcctacaaat gcacacaaaa cgaaattgtt
38161 gctttacttg cttaagtgat tcaatgaata atagttaca gaatatgatt caaaatccac
38221 tattgtttta aggcccatta agaatcaatg aaaatgattt tttttatcat ttctataagg
38281 cccatgcgtt tgtgttcact agcagtgaag tgagagtgtc aaatccctaa attcctcact
38341 ccctgcgttt gtgttatctg caagtcggag aagaagcaga agaaatcgac taatctctcg
38401 ccaatgtcgt ctccagagaa gaagaggaag aagaatcaga aatcgtcccc aaatccgtca
38461 cttccagaag atttgatagt gagcatcttа gcccgcgtct cgagatcgta ttacacaaat
38521 ctctccgtag tctcaaaaac cttccgatca atccttacтt caccggagct ttacaaaacc
38581 cgaaccctct taggcaaacc cgagactttt ctctatgtct gcctacgctt ccctgatgaa
38641 gctaatcctc gttggttgat tctctaccgg aagcctaatc aaaccctaac taagaagaag
38701 aagaagaaag aggattcaag tgttcatctt ttggctccaa ttccaattct caattctcca
38761 actgtggaat ggtcaagtct agtcgctgtt ggttcttatc tgtacgccat taccgcagac
38821 atcaaagatt ctccatgctc taacgtttgg tacctttgact gtcgaacgca cacttggctt
38881 gattctccaa gattgcgact ggctcacatc aatagtgact ttaatgggag gacgtattcc
38941 ccaggaagct ctgaaaagcc agattctttg aactgtgttg aagtgtacaa cacaaatact
39001 cagacttgga atcccgtgcc tcccccaaaa cgaaaactca aattcggaaa tatggaagga
39061 aagatttaca taccccctg tcaagagaat cataggaagg ctgtggctct aaacccaaag
39121 gttttgacat gggaagcggt aggattgggt acgaatttgg ataggggttc tttttgtatg
39181 ataggggaaca tagcctattg ttatgaccct agtgggaagt ttagatggat gaattgtaac
39241 actgcggaag gagactggag cagattagag tgatttagaag ggctacctaa gtttgctaga
39301 tatagtactg ttaaattagc ggaatatggc gggaagctgg ttgttttgtg ggacaagtat
39361 gtagctgcaa gtgggtataa ggagaagatg atttggtgtg cagagatttc gcttgaaaag
39421 cgcaacagtg aagagatttg ggggaaggtc gagtggtttg atgcggttct tacagtccct

NCBI Sequence Viewer

```
39481 aagtcttata aattcgtctg tgctaaatct gctactgttt gattgatgac tcaggacata
39541 ggtacgaagc ttcttttatt tagtatgtca cagttgaaag aatgtcttgt atcttttttcg
39601 gaatgcagtt gtcattgttt tcgccttaca aacacctgca ctcacttgct tgtagatgat
39661 taagcttatt ttgtgactgt ctctttgata attcctttga tgccaaactc tttctcctac
39721 tgttttttgc tctgaatgtt gtgttgcaaa atgcctatat ttatgtgact atgatgaatg
39781 ctctctcat gcttttagcg gatattgttt cacttctaaa catatgcact cacttgtaga
39841 ctgtttaagc ttactttgtc tatgagaata ctttgtttg tcttgcatag aaaagtactg
39901 ttatctttac atttcacgat aagagctaga acaagataca cagttagtgc ttgttttgtt
39961 gtttaaactg aggcttatgt tcatgtcttg tacttgtagc caacctaacc tacaaagaaa
40021 ttgctactgt aacaatttga ggatcttgtg tgaaaatacg ggtggacttg tcaagccact
40081 ctttctttc ttctcaatcg aacactatta ttatccaatt atatactctc tatttttat
40141 gattcaattc aagttagatt ccattatgca cagaaaatta aattgatcat atgtccaaag
40201 ctctactgtc atttttgaat tttgttgaaa ttttgaaggg ttaacttatt ttgtgccatt
40261 gaccactatg atgcatttga aagacttttt ttttgttgtt gaacatacgt gattactagc
40321 ttcataggaa gaaataagtg taacaaaata tatatatttg acgaatttct ggtcctctac
40381 tggtaatttt gagttgtgtt gtagaaattt gtggtgactc aaaactattg ttttaaggct
40441 cattaagatt caatacaaat gaattagata attctataag gcccattaag attcaatgca
40501 aattatttag ataatttcta aaggcccact aagattcaat gaaaatctgc aagtcggagt
40561 agaagcagaa gaaatcgact aatctcgcca atgtcgtctc cagagaggaa gaggaagaaa
40621 cggtcgcttg aaccgtcacc ggaatcgacc ccaaacccgt cacttcccga tgatttgatt
40681 gtgagcatct tagcccgagt ctctagattg tactacccaa ttctctcgtt agtctccaag
40741 agctctcgaa ccctcgttac ttcgccggag ctatacaaga ctcgatcttt cttcaaccga
40801 acggagagtt gtctctatgt atgcttagat ttccctcctg accctaacc acgttggttt
40861 acactctacc gaaaacctaa ccaaaacatc accgagaaga caaagaactc aagcgggttt
40921 gttttagccc caatcccaaa tcaccattca catagttcga gtatccgttgc tattgggttt
40981 aacatctatg ccattggtgg atccattgag aatgcacctt cttctaaagt ctcaatttta
41041 gactgcaggt ctcatacatg gcacgaggca ccaagcatga ggatgaagcg gaattatcca
41101 gctgctaatg ttgttgatgg gaagatttat gtagcaggag gcttggaaga gtttgattca
41161 tctaaatgga tggaggtttt cgatataaag actcaaactt gggagtttgt gttaagccct
41221 ttagcggaga gatttatata taggagcttg gtgattgaag gagaaatcta catttttggg
41281 gataaggttg tgacttacaa gcctaaagaa gatagatggg gaggtgtagg agaacatcag
41341 agtatggatt tgggattgtt ttttcattct tattgcgtga tagataatgt actgtattgt
41401 tatcgtcctg gtggaatcaa atggtatgaa tcagagaaaa gatcgtggag gaagttgagg
41461 ggtttgaagg gattgtctaa gcttgctagt agctgtgtta agttagcgga ttatggtgga
41521 aagatggctc ttttgtggga caagtatatc ccttgtagtg ggaataagag tcattcgatt
41581 tcgtgtgcgg tggtttccct tgagagatgc aagaatcagg gtattagggg aaaggttgag
41641 tggtttgatg atatgcttac ggtacctagc tcttataact ttgtgggtgc tcttgctgct
41701 actctttgat gatgattgcc atacgaggca cacacatttg attgttctagc tagcttacag
41761 ataattgctt tctatgttgt aataacttaa tgatgggatt cttcttacca tattatttca
41821 aaatgcaact ggcttcattt ccttcttcat ttttgttttc ttgtcaagat caaacaaaga
41881 ttattcagtt atagcgtctt ttgttgtcta gaaatcactg aactagtagt tagttattac
41941 accttcaccg gttgaaacaa caacttgatg ggacagggaa gtgagaatct tttttgcctt
42001 ctcggtctta gaaagaatgt acatttgcga tgagagtaag agcaagaatt taaaaaagaa
42061 caaagaggaa aagccgaaga aattgacatt gttttgcttt ataaacattt gcagtcactt
42121 gcttgttgat taagttgatt ttgaggttaa ttgacaacta taacaaatca gaagatttaa
42181 gttgcaatta tgtgcgctaa gctaaggctt ttagactctt atcatcatgt agtaacatca
42241 agccaagtgc caatacctttt atatagatac ttcctgaata caacaaaatt ggacacaatg
42301 atctgagttt catattatat aatataaatt aactaaaccg atttctcaac cacacatttg
42361 ctaaaggctc aatataaatt aattaaccct aaatccttct tcttcgacct tttatttggc
42421 gccaaacccg cagttacatg tttttatttc ccgctcgagc aaattgatct gatccgagtc
42481 tttcaagcgg ggagagaaag agagagattc gttagtacat tgatctaatg gcgtcgaatt
42541 cggaagcggg aggatcggcg gattacgaaa cattgatgtc gacatcggac gttgagctat
42601 tgaagagggc gtggcgtaac gagaaggcgg cgccggagat tcttcagtac gaaggagctc
42661 ttgtcgatag agccaaagaa cagatcgaat tagtggtaaa atgtctaaaa ccctaaatta
42721 ttctgttgga ttttttccca gaaattaacc actaaatcag aaatcttttg gtctatgatc
42781 caaattagac tttccaattg ttagaatagt attcgttttt gtgtgaaatc gaatgccaca
42841 ttaaacttag ttagagttgt tcattgtctt gtcactttct agtggttgca tttggaaatt
42901 atcagttttt gaaatgtttg taaataaaca ggaagaaaca attgaagatt atgtggaaaa
42961 tggcatagat cctctggtgg tgtctctata tcagatggat ttggacagag cccagtttct
43021 actcagatca tatctcagag ttaggcttct caaggttttt ttttttttcc aagacttcca
43081 tattagtttt gcgagtttat caacgttgat tccaagtgtc aatgttgctt aattgacttt
43141 gcttttatgt catcttccaat ggtgttcttt tgattgctgc agatagagaa gttttatgttc
43201 cacaacctca actcagaaga agctgaaagg cgacttttctg agcaagagaa agtgttttgct
43261 accaggtttg ttttcttact tgtctgactt tatgttaatt ggctgtcttc ttccttgtat
```

NCBI Sequence Viewer

43321 gcgttttgtt ctaaatgtta taggtttggt gtagaatctg tgtggaatag atagtgttta
43381 gagatttgtc tggtactact ctctacatgt ttgagttgtg aatgggatac aatgaactta
43441 catattaata tagttcttgt tcatggagta tttgaactgt cccgagaatt gaaaaacttc
43501 aggtagccta ggcacttttt tcatgaatat tttgtttgag tggtttttc agactttaaa
43561 ctcaattttt ttacaattta cgttagtggc tgctgttgtg agacatatat ttgaatgatc
43621 acaggtgtgc tgatgatttg gcaaagcatt tcgaagagac tgtactgctt aagctgcccg
43681 agaattatca gtcggttctc aagcaatccc ttataagtga agtggatgat atgggtacct
43741 ccctaaatct tattactttg gaccataatc tcagttggtt aagattgttt gctattatat
43801 ctgagcattc ccttcacagt agctagagtt gtgtttgtgt tagtgatgac tacacttgtg
43861 aacttacagt gccgcagccg catttggata cctttgttgt ttgtagaagc aagaactttg
43921 tctccctcaa tctatacgaa gagtgagcat tctcttattc ctccatacac atatatatat
43981 acacaatctt tctgtaatct aagtgtgggt ttgcctttac ttaccttggc agaggagaga
44041 gccccgagac tgtggagatg gagcgcggtg atctctactt catacgttac aagattgtaa
44101 aaagggcaat cgaatctggc caaattgact tgatctgagc tcgccttttt tttcactcaa
44161 atctcttctc atagaataag tttcggatga tgctaccagt attcactctg tatgatcttg
44221 tgttacttga gttgtatagt gagtttgagt ttctgttctt gatataaact tagtagttag
44281 tatctttgta gcttcttgca gttgtgccta gcttagctat ccattggctt aaccataacc
44341 agtggcggag ccagaacaat gactcactgg ggtcaaaaaa aagttaaacg aaacgataaa
44401 ctttactata aacataacaa tgtcattaca aataaaaact ctaaagatga actctacgtt
44461 cactcatatc ttggaatcgt tttatcacga ttcgtttgta acagtgttaa gtacttgctt
44521 ttcaataaaa caaacaagac aatcacttaa gaacacatca ctgatacggt tacgcaatgt
44581 agtcttcaca aaattcatcg cggaaaaaca tctttcgaca gttgccgttg caacaggaag
44641 aatcaaagat agctttagaa gacgatatac caaaggatgt gaaagatgct tttttgtctc
44701 taccatcaca cgagcaagat caccaagaca ctttaaatct gtaaactttt catcgttttt
44761 cacattatca aggtagatct caagttgatg atcaagagat ctacgctcca caaaactaaa
44821 gtcatctgga taaaactcga tcaatctcac gagcatggac ttatcaaact aacgaaaaga
44881 atctatcgga ctcacagaat ctatggtcca taactaaacc acatacttct ttttaaaagg
44941 gcctgtctgt ttattttta ttttcttaga tctttttgta attggtccat gatgactgag
45001 taagacccca aaagccgtaa aaagtaaaaa aagatggaat ctttatagtt taaaaacaaa
45061 aggctagtac taaaatatct agttacggtt acgtgtgccg ccatagaaaa acaaatcttc
45121 tgctcagtct tggattgaga ggtaagagat ccccaaaact tcgcttttca atctctgaag
45181 ctgcttttga attctcatgt tgatttcttt gaatttgcat cgatcaagtt taggtttttt
45241 ggatttgggt ttctcctaaa tatcttcttt gaccagtgtg tgtgtggttt ggatttatcc
45301 ttttggatct ttgtttctcc agaataagta cagaaattgt aatctttgtc aaattttgtt
45361 tattttgagt ttcttagtga actcagttgg ataacagtga cacatatgag atttgagatg
45421 aatttatttt ctggtaaatg catagagatt attaatgttt cttactgatt tgagaattgc
45481 aaggcattgt cctttgagtt tggtgttaaa gttagcaaat tggtgactat ctctattgag
45541 tttacatgtt ttcttggttg ttattgtgag gatttaagat ttgcctgctt ttgtatttct
45601 ctttggtctt gtaatgtcag aattgtttct ttcttaaagt tgatttcttt gtcaatcttt
45661 aaacagagat cgaggtgttt ttgacgcaat agaagatcgt tctgcgtgtt attatccaag
45721 aactgcttta gtgttataaa catttgagca gagatgcatt tgtggccatc gttgaaactt
45781 cgaaactctt tcaagagcac tagccagaag agatatcaaa ggatgaacag tgggaaacaa
45841 agtcaaagca atcagcagaa actccttgac ggcgacaaag aatctggaag ccccgagatt
45901 tcgtctggtg gtgggtttgc tgttgtttgt agagatgttg ccatggttct ttcttgctgc
45961 tactgctgtt tctgctgtgg aggtaaataa accttctcga cacttttccc cttagctagt
46021 atctgtgtga tctctggacc aatgttttcg tatgaccttt gagcattata tgcaagcacc
46081 atggcacatt ctcatcacat attaacttag ttggtatcag tttagaatct agattaaacc
46141 atgattttta tcttttagca gatatacaat ggttttcctt tagtgattat tcgttcttct
46201 tcatcaattt ctccgattct gattctgatt cctctgattt gttcagcttg cattgatgag
46261 gaggagattt gatgcaatca tgtaagctcc tggtggagga gttaagtgaa agattttagc
46321 cttagaaaac atcaatctct caaaatgtaa ttagaaatgt tgtttataat taatctgaat
46381 tctttttcca cctttcccaa atcaaattta tatattatga ttgaaattca cgatattgac
46441 tttatcaaat agataagcat ggttaaatga tttcaaatga ttttctattt aacacaaact
46501 aaacttaaac agacacaact tttgttcatg tagaatttta actattcatg gcatttagga
46561 tttgaactgt tgaatttccg tttaaaatta aatgtcattt ttaatgcatt attttaaaaa
46621 ttatttacat ttttatgact ttaatttaaa tgtattttaa taagaaaatt tataaaatat
46681 attaataaat agtaaaaaat aggtttttca tatgtatgtc aaattttaaa attgtttttt
46741 ttatataaac gatatttatt tataaacgga aggaatataa ttttatgagt tttttttctt
46801 tttttttttg tcatccaaaa ttaattaaaa ctcaaacgag ccaatctcga gggatattgt
46861 ttacacatgt gatatgatgc ggttcggtac ggattttccg cgccaaatta ttcgctttta
46921 cattagcact tcgagaaatt aaagataaag aaaaatttgt gagttcttcc ttgtcgcttt
46981 gcaattcctc caaatacaca gaaaatgccg gccattcggt tggacaattt tatgaattat
47041 gagttagaaa tttagaatat aattctctat ttaggaattt agaagattgt tagaagagca
47101 aaagtcaaaa ttactaagtg agagtattta ctcgactcag ccaaccaata ataatccgac

NCBI Sequence Viewer

```
47161 acgtcatata aaagttccag aaaggtcaaa aaaaatcaca cacatttaaag tgctttgaat
47221 acaaaagcag actccactat ccaattaaat tacgaaaata ccacttcggt taactccact
47281 cttttcatct tcttctctgc cgtcgtcgct cggagctcgc gaagctctct catcagaaat
47341 caaaaaccct aaaaaaagtc ctttctgatt catctataca gacacatact gtattctcaa
47401 aacaatggag attccttctc tgaataagca gcaagagttc actctcgcct ccgtcactga
47461 tctcacttct ccttcctcat ctctttcgtc ctctcctgtt gtagctacgt tttcatgcgt
47521 taatgaagtc aaggaactcc gatttcagga atcgaaatcg tctgatggtt tcagttttga
47581 tcttagttcg actcaggtgg ctctcgattc gatttacaat ttgcacttct tctggttaaa
47641 gaatctactt taatctcgtt ttgaaacgtt tctatgcttg gttaagtaat gcactaaaac
47701 aatcttctgt tagttaggat tattaaatta tacgtgtggt aagtttctgt tctttgagct
47761 agattctgat tagtgtattt gattttgtgc tgctcagttg ttcaagttgg ggccacttca
47821 gttcacttgt gtatctgatg gttcgatttc ttctgcaaag gaggtacgtg aaacatttct
47881 tttgttcagt ttctttctgag tgttcatggg attggtttag tgtagttggt tgcaggagat
47941 gacttgatcg actgcgcatc cttgtttact atatctagtg tgtaatcgga aatgttttg
48001 ctgattttgt taccagaaat cgtcattttc tcgaggagtt gtaataaaat ttagagatga
48061 gaaggatagc aaggagtttt gtgattcatt tgaggaatgt aaaaaggatg atgctgttaa
48121 acaaggtgcg aatcaatgtg ctaattcccc atcctagttg tatcgttcat caaaatccca
48181 tttcattgat cataatatat tcatgactct tgatcgtact cattctttga tgtcttttttc
48241 gttcttttgt ttctaggatc agccttgccg aacggaacag ttgtttcagc taataaaagc
48301 aagttcgatg ataagattga ggctgcttca gccaaaatgt atttccatta ttatggacaa
48361 cttttacatc agcaaaatat gctacaggat tatgtgagga cgggtaagat ctggcctttg
48421 tgcattgatt acctcagagg aaggataact ttttttttgt tgttgcgaac agcaatttga
48481 cccaagatag ctcttttttcg tcatattact ctcttcaggt acatatcatg ctgcagtcat
48541 ggagaaccgt tcagatttct ctggccgtgt tgttgtcgat gtaggcgctg gtagtggaat
48601 tttgtcaatg tttgctgcac tggtatgcat aaatttaaga tattgtgggc ctagaggtct
48661 actactacta ttcgtctgca ttttgctttc actttttctg ataaacgtcc attatttttct
48721 tcctaattta ggctggtgcc aagcatgtgt atgctgtgga ggcgtcagaa atggctgaat
48781 atgcccgtaa actgattgct ggaaacccat tgcttgctga gaggatcaca gtaagacctg
48841 tagtccagtt tcgttttgtaa ttttttttcat ttatatccat tttctgatg ttcttcaact
48901 tgacacaatt gtatttgtag gtcatcaagg gaaaaattga ggatattgag ttgcctgaga
48961 aggccgatgt tttgatctct gaaccaatgg gtacgtgttt cagagacgga ttcttctttt
49021 cacgtgctct tcatgaaaat cagaatcttt cagaaggcaa attttggtct gtgagaaaat
49081 aaatggagaa acaccattgc tgattcaaag ttatgtatga aagaatatct cttatatctg
49141 atatacctgt gtctgtaggc acctcatgg tcaatgagag gatgctggaa acatatgtta
49201 ttgctaggga tcgttttctg tctccaaacg gaaaaatgtt tccgacggtt ggaaggtaaa
49261 gctgatcttc gatagtttta cttttctatc tcataaatgt ctttttgtc atgtaacttt
49321 gtcactgtag gatccacatg gcacctttcg ccgatgaatt cttatttgtt gaaatggcaa
49381 ataaggtaca tattgcacat cttatcagga agatatgttt gagattgctt gtaacataaa
49441 ataaaatggt tgaattaagt atgggatggt taggcttcca gtaacttagg ctttctcacc
49501 tctctatgca ggctttgttt tggcaacaac agaactatta tggagttgat ttgacacctc
49561 tatatgtatc agcccaccag ggttattttt cccaggtata tgcatttaca accagtgaat
49621 cactgatatt tctttgcttg tctcatcctt tacaattttt catatttact gattttgtag
49681 cccgtggttg atgcatttga tccgagatta ttggtggctc cctcgatgtt tcatgtgata
49741 gatttcacta tgatgacggt aaagctattt cctacattag attcggttcc acggtgtttc
49801 cttaaaattg aggacacttt tctgtcatgc tatttctatg tgcagtcgtc cctgttagt
49861 gaatggcatg atccaaattc aataactgtc aaatttttttt cacaggaaga acaatttttat
49921 gagattgata tcccgttgaa gttcactgcg tctgtgtgta ccagaattca tggacttgcg
49981 tgctggtttg atgttctctt tgacgggagg tatcaagccg taacaatttа tatgttttca
50041 tcctgtgcta gctaaagatt aatattcctt tttgaccatc aaagcctcaa aggtatatca
50101 tacaaacctg atgctttgta tccaacagca cggtgcaaag atggtttaca actgctcccg
50161 gtgcaccaac aacccattgg taccaaataa gatgtgttct ctcgcagcct attcacgtca
50221 tggcagggca agagatcact ggtagacttc atttgattgc ccacagtgct cagagttaca
50281 ctataaatct aacactctca ggtcaatttt gctttctgac atctctcctt ttgaatcata
50341 tcatctactt tggacttatt ttctgattgg ctttttaaac acagctaaaa tgtggggtcc
50401 cggtgccaat caaggtggaa tcctccagac atcatcgtgc aaactcgatc tgaaggaacc
50461 ctattataga atgtctcagc cacaggttta ccctacacag gaaccaccag ctcagtcgca
50521 agtaagcgtt tgttattgtc ttgattaaac aatcaggata tgattgctgt tgtgtttaat
50581 catttggcgc tttaaagcca ttacattctt tgaatttgca tatttgcaaa agtatatatc
50641 agcattaaac atatgtggac attctcttgg tgcaggacat acatatacat agtgatgact
50701 tggaagaact agagttgcta caacaaaacg caaacgctca gctctagaac tagaactaga
50761 aatagaaccg ttgtttctgt ccatgtattt gtactcttgcg tctgcttcct acggtttagc
50821 accaaccctt tattttcaga cattatcttg cgttttcttg tctagttcga ctctatttca
50881 ctacgtatca cccacatggt atttgatcct ctacacgcaa atgaatcatt ctatttggg
50941 cgctaagcca ggttaattga acttctgtac aactatgttc tgacactgca atttttaaaa
```

NCBI Sequence Viewer

51001 ttgatttgga gttgttttca aatctgcaat agggacctag ataattcata actggagttg
51061 acttcaattt attgagttta ctgaaacaaa aatgaatcgg agaacgtgca aattgtgcaa
51121 tggacaaagg aactgtaata gtatagagca aattttaggc taagctgttc acagctgcaa
51181 ctgcaggctc aggtggatta gcaacaatga cactgaagca gcgtccacag agagtaggat
51241 ggtcagaaaa tgatccgacc tggcccgagt aattccagca cctctcacac ttagacccct
51301 cggctcgtga caccccaatc cacactttgt tctctcctc cacatactct cctgtgtgtt
51361 gcacactgct tatcatctcc ttctccattg aagacaatac ttcaacctgt aatatgcatc
51421 cacatcagtt tcatggctat agtctcatag aattacactc agcatgatca aatctcaact
51481 tcaatcgaaa attgtggata ataattgagg aatttgagat gaatcaccat caccaagtat
51541 aactaaaaac acaatatatt gtaattcctt catacctatt tcttaggagt ttaaaggggt
51601 tttcgagagt gacttctata gatctgatca gtacatgtgt ctactgctac aaagtactct
51661 tcttttaaca tctaaacatt tagcatttct ctgggggtcc aaatactata tctaggataa
51721 agaaccagac tggattaaaa caaattccgg tttgtgccta atcctattca cactaaatct
51781 tttacttatc tgttttggga agagcctact tgtagtttgg cctgaacttt acacagaaaa
51841 accacggttc acataaaaac ggccctactt gttgatggtc accatgattc taatttatct
51901 taagctataa caggaatatg aattttccag ataaggagat ttagtgctta cttgtgatgt
51961 tatgaatatg cgttgcagag tgtcagcctc gttctttgct tcactcatct ctagtaattt
52021 tgcagccatg cctgcatctg cagtatgcaa atacacctt gcttctaaac tggaaccaat
52081 cattttctca ttacgtgcta gctccagcac tttgttcacc tcagtcctca actgtcacac
52141 aaacaccata agtatttcta agatcctcat tacaaaagc gcatcaagaa taaggccagt
52201 aaaacaaaga ataaacgtaa aagttcttcc agagaaattt cgataatgcc atatcataag
52261 aactgattac acgatcaact tgacataaca tcgagttgaa tattctgata ttgtgtaaga
52321 tttcattgag ttcatctcca ctaagtagct aaggaactct attactttaa ttacaagcag
52381 agtagagagt atacaatggg ctgaattgct gagttttcc gaattgattt aggagagttt
52441 tgatacctcg agaagccttt gccagaagag aacatcttca gcaggaaatg aaagccattg
52501 ttcattcaaa gtaggccatt taagttcgaa gacaaattct gctgcagagc catcttcatt
52561 tctgtactca aatggaagat tctgccagac atcttccgct aggtgaggta cgattggagc
52621 aatcactctc aatatggata aaagatgtgt tgaaagaact gtttgacagc ttctttctggt
52681 aaaacttgaa gtaccccgta gaaaaaagat ttttggtcca gattacgttg aatagcacag
52741 acgtattaag aacaggtcta catagttct agcaaatatc aaccaaaaga agatgagaaa
52801 cttacccagt atacagtcta tctttagcta tatcgaagta gaaatttgac agatcaacaa
52861 ttgtgaatcg ttgtatgatc tgaaataaag aaagctcatt tagtaacaca gatgatctag
52921 aacatgaaat taagcctcat gaggttcaaa ggtccattgc aaatacatca acctcactct
52981 caccttaatg gatagtgatc caattgctga actaaagttt cttagatata ttaaagcttt
53041 taactgataa attaactcct gtaagagtga tctatgtctg tttcccgtca caggaaaatc
53101 aatacatacg tgaatgatgt accaaaataa gatggagagg ttacctggaa tatttaaag
53161 aactggtagt tctcgtaaca ttcttgtata ttctttacaa cattctcaag ctgaaagaga
53221 gcatgctgat caatgatagg taaatcctgg tatggaacag catttatcaac ctacaaacca
53281 acacaagtta aatattaaca gattgtcacc aaccccagaa tataaaacag aagcaagaag
53341 tcttgacaag tgggaaaaga ggtgaaaaac actactctat gtaaagctta agagttagaa
53401 aacctactct ccaatcatga agatttccca agaggtatct taacgttccc cttaacttcc
53461 tatatatatc agacatttga cgaagaatct ggggacctat tagtacatcc cctgttgtaat
53521 ctacactaga aacccaaaga cgcataaccat cagctccgta agcaggtgcg tcctgttgta
53581 agaagacaaa cccatgtcag gaaacagatt aaatacgaa gatagttcta tacaatggga
53641 tttacaatgt gacaaacctt tgaatttttc cctccttcaa tgaccaaacg tgggtccact
53701 acgttaccta gagacttgct catttcata ccctctcat ccaatacaaa tccatgtgtt
53761 atatacagcag aatagggggc ttttcctata tacatcatg caaaatgtga agaaaaaaat
53821 caaaagccac acaaagatag ctaaaaatct gtagaataat ctcaatttat aaggcaatat
53881 catggcatgc acttctcacc ttgcgttgcg atgcttgtta acaaagaact ttggaaccat
53941 ccacgatgct gatctgtacc ttctagatac acgtctgcag ggaaactaag accttcgcgt
54001 ttccctaata caccagccca agaagatcct ggacatacag atgggaaaga aagtgatgta
54061 aaataggtgt tgggtgctta gtgtgactat ctgattttaa ttagcttctt ttatcatagt
54121 acctgagtca aaccagacgt ccatcgtatc agtcccctt tcataatcag ctgcttttatc
54181 acgatatttc tcaggaagta ggtctttccac tgacatatac caccacgcat cactaccctt
54241 ttgggaaatt atagctgcca gacaaggtat tcaaataagg ttacagtaaa cacagccagt
54301 aaataacaac caataaaaca caccatactc tcttataggc ccttacattt aacatggttg
54361 attgtctctt cattcatgag tggttctttt gtcttaacat gataaaaagc aggaatagggg
54421 acaccccatg tcctttgtct tgagatgcac caatcagatc gactggaagt catggcagat
54481 attctgttta ctgcctgaat gacaattgcc aaaagataac aagtaccaaa attaaactcc
54541 aaacaatata tatgcatctc ttaaaaaaga taaagtaagg aaagacgcta agtcaagcag
54601 tacggtatca tctcaacatt gaaaaaatct cacagcaagg tggatcagat catgtggttc
54661 atgtgctatg aatctccttt cacagctaac acccacacat gatacttttа gcccacataa
54721 tgaagaaaaa aaaatcaact tataaactca gtatgtattc aactaggtac ctgatgtggt
54781 acccatttta cattgttgat tgcatccatt gtggctgtgc gaaacccttc aaccgatgca

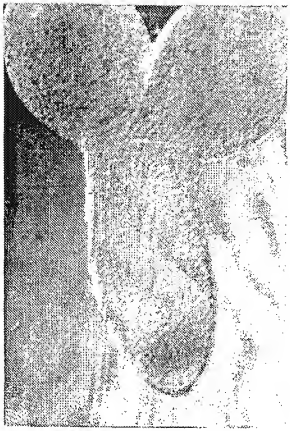
NCBI Sequence Viewer

54841 aaccactgct cagtcgctct aaatatcgtc ggtttcttag ttcgccaatc atatgggtat
54901 ttatgagctg catgacaaga aggaaattaa atcatgagac tctatagcaa atgtttcctc
54961 aaatggttcc tcagaaagca tttgaatcaa actcaccata agattcttcc atgaccagag
55021 acatattttc atccaggtaa ctgacaacag cggtattccc ttctccaagg acagagagcc
55081 ctctaaactg tccagcttct tcagtgaact ttccttcatc atccaccgga gagacaagag
55141 gcagcccgta cttcaggcca gttgcataat cctcctgacc atgaccaggg gcagtgtgga
55201 ccaaccctgt tccagattct gtagttatgt aatcaccccc tataacgact ggacagtctc
55261 tgttgtcaat gggatgagtg tacctacaac catgatgaca caccatagac taaggggttc
55321 aatttatagc attacatgca aaactggaat ttctttttatt ctctaactta aaatgtaacc
55381 tgcagttttc aagatctgaa ccaaggaatg ttttacttat gctaagctte acaccccatt
55441 tcgcttctaa tgctggcaca aggtcggtcg ctacaatcac gaaaagcttt tgttgattct
55501 ttagaacctt cccgggaatt tcttttttat tgcttgtcac tgttgattca tctttctgaga
55561 atgactgcac ttccacaaca gagtactgaa gctttgcatt caccgcgaca gctgaaagtt
55621 agacaagaaa gaatagaaga agagtcactа aatgtgctta tatagcattc aaaaaaggta
55681 cgtattattg cttctaagtg gagaagagat atgagaaaga acggatacaa agtaaagtag
55741 gatgcaattt gacagctgag agaaacacat atactaatat aacccctтac cagcattggc
55801 tggcatagtc caaggtgtag ttgtccatac tgccaagtat atattaggga taaactcatc
55861 taaaaggctt gttttcgctc caccaaccaa tttaaaaata gcatatatgc ttttggaaat
55921 atgaccctcc ggatactaca cataatggca gatacatgtg ggactgatat cagatаacaa
55981 ggtaaagtgt ggaattgtca aaatattatt tcatgaaagc ggcatgcgtt gagaacagca
56041 atagaaaact acattaaatc agggttcatt ggtgatacat aaagaattag tcttcgacta
56101 atggacaaca gatcgcatca tggtaaaatg tcaacggaag taagaaagag aataggaaga
56161 ctggtatgac tttagcagac ttaagttttg cattgaactg gccgtaatat tattattatt
56221 cttagattat tgattcaaat ttacctctag ttcagcttct gcaagagcag tccgagatga
56281 agggctccag tgaactggtt tcctacctct ataaatgtac cctttcaagg ccatctggcc
56341 aaatacttca atctacaata tatgataaaa atatatgttg actttatcag gctttgaaga
56401 actctaataa tcatctccta cacggttatt tcatccagta aataatgtgg gacgtgagac
56461 cattacctga gctgcttcat attccggatc aagagttaga taaggattgt tccagtctgc
56521 ccatactcca aatcgctgta aacaccaaga ataatgtcat atgtgctaca tgtcaatcga
56581 atacccaacg aaatatcttt cttaattatt atgtcctgag agatataaga atcattatga
56641 cggttgcatt caaaacaaga aatcagatat gaaaacaaag cttgatagcc accttaaatg
56701 attccatttg tgttttgact gttgcttttg caaatttcgc cgccttgcc cttaatttta
56761 atggtgtaag ttccttttctt acttcctgat ccagggactg cagaactgaa acgtttgacg
56821 gttacaccaa attacacagg caagtaagta actggaaaat aacaaaagca cagttacttt
56881 ctgtttctaa tgactcaaag gtccataacc tattgtgtgt ttaccttca actcaattgg
56941 aaggccatgg caatcccacc caggaacgta ctgaacttta tagtttgga gcagctgcat
57001 gaagaagcag gcagtttgat gagaaggaac tccactattg gacgtaagga aaataaagaa
57061 tccctaagga gagtaaatgg atgtccaagt aggaaagaca aatacctтat agcgattaat
57121 gatatccttc agtatcttgt taagagcatg acccatatgc agatcaccat tggcataagg
57181 agagccatca tgaagaatga aactccccтg aaattctaac taactatcag acaccagtct
57241 ctcaaagcaa agttggcata agataatttt gagaaactca ctccattgtt attatcagaa
57301 actcttctga agacctggtt ctcttcccac aacttctgga gttcaggttc ccttgtcaaa
57361 gaattcgctc tcatacaaaa accagtcttt ggcaaatcaa ccgtttgctt gtacttttcа
57421 tcttcttgct tttccctaa ccaacacaca acggtaaatg agtgccacat tgcaatatag
57481 tattaagaaa gaagtacaag caatgtgatg caaaccttca gatgctttct tggctgccat
57541 aacaggccca cgagaccttc gttttgatga atgaccaaac tcattgtttg gttcagttga
57601 aaaccgtgag acatataaga atgaaaaagc cttgacggaa gaccttccct tggcaagaaa
57661 agaatccaac ggcgtgtttc tccttaaatt tgaacaactt ttacctgaca aaacctaaac
57721 ggggacaaac tacatctcaa aaactatttc cacgcaaatt ttcagcaaga cacaataact
57781 tctatcagat tgatataggt ggtagaagta gtttcctccaa acaagttcct gagacaacac
57841 atttttctaa gaaagaaaat agagcataga aacagactta gtttggttaa taagggcttt
57901 aagcattaca gaggacacat atgagctcgt catgtgatgc tcagatagtc agaactcaag
57961 taaactactt gcaatttcat tcttcaaaga tataagcaat ggataaaagt caaccaagaa
58021 gcctctgaaa ttttatcgaa caactggtgt gcaatagtaa accttaagaa cacaaacaat
58081 aacagaaagc gaactacttt aacgaaaccc acaaaccgaa attgccaatg cattgataat
58141 aagaaccaat tcgtcgacaa atgtgttcta ggaatcgaac ttgattagaa agactatcac
58201 tttccccagc acacaaaaca aagagtcaac agagcagaga tggattcaaa gtaatcaaat
58261 ttacattcac atactctgta agatgaagat tgaaccatcg ccattgcagc agcttccctg
58321 ggatttcccg cgaaggattt gaagaaagac gacatctgct tcaagtttct ctctgtcgct
58381 tctatctatc agagctcatt ttcacggcga aacaaagagt gagaaggaag gaagaagact
58441 aacgtcacca aattcaggga agcgtatgtg tgtatgtgtg tggataactg aaaagttgaa
58501 actaaaacca aacctcagaa ctcggtttca atttttggat tttatccgtt aaatcccctaa
58561 atctcttgtt ttttttcttc aagctcatаa tagtaaaaga tatcacgttt taaatacata
58621 aatatactaa caagtaacaa gaggcgaaat tcttttaaca tgtatggtaa atagattaaa

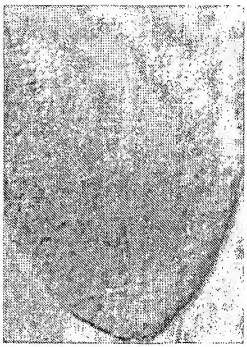
NCBI Sequence Viewer

```
58681 tataaacaaa aacaaaaaca accattagat aaggcgggtg aatttatagg catatatcat
58741 cagtaatgca agatataaca gttagactca acagtagcat caccagtctt catatcgaac
58801 aacttggtcc gtgcttcgac ataacctcta gcaaaccgga acattccact gcctccaacc
58861 accggcattt ccctaacctt ggacatcacc gtgttccgac caagaatcgt tatggtactc
58921 ccgttgtact tcccagtctt aaaagccaag ttcatcgcca tcaacaaccc tatctccttt
58981 tgggctgctc caacgtacat cccttgggct tgacccacca cagtcgagtt aatgggcaca
59041 tctaatgtta aagcgttgtc tatcatggtg attgatccga aatatgagga acttgagtat
59101 tttgcaaccg gcccttggat cctgatggaa cttgggtttc gaccagttac aatgtcgtgc
59161 caatagacac ggagatgtgt gagtttctct tttttgccga ggccgaggtg ttttcggttt
59221 attgttcttg cgaagtcttc accgtttttt ccggctaagg cggttgcggt gaggaggagg
59281 atttgtgcgg tgaggatgag aatgagcttt gacatgttag ttttctggtg tcgtacttta
59341 attgttttat ttgttgttga tagaatgata gtttatgtaa attgtgggtg tatatactgc
59401 aactgtgcaa cgtgctatat atatcgagta atagatgaga tcacgagtgt gaaaattaaa
59461 aaggggaaat agataaacga gtgtctacat atagaagaag acgtgctatt tgataaatct
59521 tctagaaaat aaagtgagtt tggattattt gattttgtta ggtgaataaa gtgtaaaata
59581 ttgtatttta tgggggaagt tacaaaagat tcaatcaata tcagtttcta aaagtttgat
59641 tggcaaagta tttagatgag tatagctagg tgagcatttg taagttatac gattcctcta
59701 agctttgatt tcgagtcttt tgttgggggtg ttaaattctt aaaaacaaaa acaatattat
59761 atcctttttt taattaaatc tcacgcttta ccttttttatt tcatgcaaaa gattgaataa
59821 ttgaactctc tccactttgg tttgtttatt catataaaac caaacagaaa ataaaaacac
59881 cttggtcata accttaatag accgatataa tcgggaaagg aaaacaaacc taactattta
59941 aatgtttctt taacaaaaag tattgaaact gacctgtttt agttggaaca atatgcctcg
60001 cgtgaatatc ggtccagaca acaacgctca aatagagatc actactctac aatcacaaag
60061 caaatcaaga aagctaaaga aaatcaaaat ggctagtgga agtgagcgtg tggtggttgt
60121 cggagtagat gattcagccc atagctacca tgcactcgag acggctcttg atctcttttt
60181 catacctttc aaggccaacc cacgattcaa acttgtcgtc cttcacgctc gaccaacagc
60241 gaccttcttc ctcggtgttg ctggccccggg aacggtcgat attataccaa tggtggaaga
60301 agattttaaac aagaccgcgg acttggttaa gaagaagtgc gccgaagtgt gctcggctca
60361 gtccgttgag atctcgtctt tggaggtgat tgaaggagat ccgagaaaca taatgttgga
60421 agccgtggag agacatcatg catgagttct tgtccttggt agccatggtt atggagctga
60481 cgtcaagagg atgttttttgt ggagcgtcaa gtgtgcgtaa ctatttagct cattaagcac
60541 attcctccgt tctgatcatc aagaagccta aggataaacc tacttctacc aataaacagc
60601 aatgaggaac aagacgacgt gattttcgta tatgttacca ttctgttttc ttaagtaaga
60661 ccctatctat cgaacattgt gcatactctt cttcccaac atcttccatt tcaacgttaa
60721 tagcttagct ctaatgtata tactcattgt atatattcac tatattaatc gcctcttctt
60781 tgtgttatta atctccttac ttttttctgt atatattgca taaaaatgaa gttcttagaa
60841 catttttaca tttatttctat atgatcataa ttgaaaggtt ccaattaaat ttgttctttg
60901 gaaacttgta gtaccttgca agagcttga tcatgtacaa agctgaggta gaaattaaac
60961 cttatcgtga atgtactata accaatgttg ccgagaactt gtaagcaatg gcttcagtaa
61021 cttcttttc ttttctcagt ttacctttac cacttgtga gtcttgttg catagtaagg
61081 tttgatccga gcttgtagaa gacgccactt gtgttttatg cgaaagacat ctaaaactcc
61141 tataaccaaa taatcatgtt gctgactgct gaggacttgt aaacaatcga ttttagcaac
61201 ttttttcttc ttttctcagt ttacctctac cattttatga gtcaagtttg catacaagg
61261 tttcaatccg agcttgtgta agatacacct tggtgtctta aatataaaaa gatccggcaa
61321 tgttgatttc cttctaagca atatttttcaa tcttcaaact catttataca tagttaacat
61381 aacctccaaa acataaatat cacagatcca aaagatcgga taaacataat ctctccaaat
61441 aatgatggct tactctcata ctctccagca gttctcaatt ctatctttca tcgtcacaat
61501 atggaccgtc caatcggtac caccacagcc accggttcga tgcgatcaaa ccggctgcac
61561 cgtctccaat gcaaaagagt gtctttccaca aattcactga gcccttcgag aaagagagac
61621 acaatgccac aactaaaaat tgaactgaaa attgtagggc tttaccccaa agggagccct
61681 tccagaaatt ctcgaaggcc caagacaagc tt
//
```

FIGURE 15



A



B

FIGURE 16

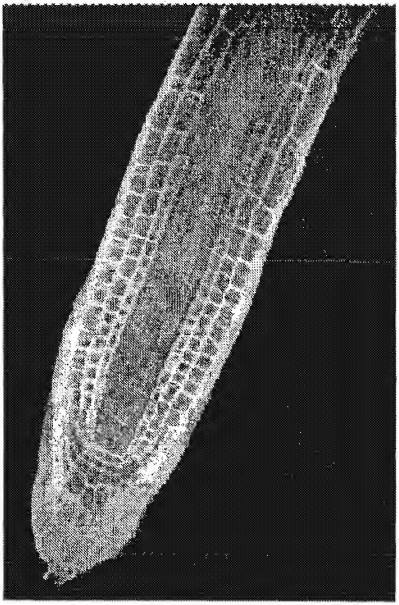


FIGURE 17

RCH1.2 AB028621+AP002037 (BAB03091)

```
mslhslilffsssssslffsfffiifcfslsdaeqnpeasilyswlhssspspsslslfnwnsidntpcnnwtfitcssqgfitdidiesv
plqlslpknlpafrslqkltisganltgtlpeslgdclglkvldlssnglvgdipwslskllrnletilnsnqltgkippdiskcskllslif
dnlltgsiptelgklsgleviriggnkeisgqipseigdcsnltvlglaetsvsgnipssIgklkklletlsiyttmisgeipsdlgncse
lvdlflyensIsgsipreigqltkleqlflwqnslvggipeeignсsnlkmidlslnllsgsipssigrlsfleefmisdnkfsgsipitti
sncsslvqlqldknqlsgfipselgtltkltlffawsnqlegsippgladctdlqaldlsrnsltgtipsgifmlrnltkllIisnslsgfipq
eignccsslvrlrlgfnritgeipsgigslkkinfldfssnrlhgkvpdeigscsseltqmidlsnnslegslpnpvssIsglqvldvsan
qfsgkipaslgrlvslnklilsknlfsgsiptslgmcsglqlidlgsnelsgeipselgdienleialnlssnrltgkipskiaslnklsil
dlshnmlegdlaplanienlvslnisynsfsgylpdnklfrqlspqdlegnkklcsstqdscfltyrkgnglqddgdastrtklrltl
allitltvvlmilgavavlrarrnidnerdselgetylkwqftpfqklnfsvdqiircIvepnvigkgcsgvvyradvdngeviavkkl
wpamvngghdektknvrdsfsaevktlgtirhknivrflgccwnrntrllmydympngslgsllherrgssldwdlryrillgaa
qglaylhhdclppivhrdikannIligldfepyiadfglaklvdegdlgrcsntvagsygyiapeygysmkiteksdvysygvv
vlevltgkqpidptvpegihlvdwvrqnrgslevldstlrsrteaeademmqvlgtallcvnsspderptmkdvaamlkeikq
ereeyakvdllllkkspppttmqeecrknemmmipaaaasskemrreerllksnntsfsasllysssssi
```

RCH1.3 AB011476 (BAB09286)

```
mprnprfcfflflfhssIffsipcfsideqglallswksqlnisgdalsswkasesnpcqwvgikcnergqvseiqlqvmdfqg
plpatnlrqlkstllsltsvnltgsipkelgdlselevldladnslsgeipvdifklkklkilslntnnlegvipselgnlvnlieltlfdnkl
ageiprtigellknleifraggnknlrgelpweignceslvtiglaetslsgrlpasignlkkvqtialytslIsgpipdeignctelqnl
yiyqnsisgsipvsmgrlkklqslllwqnnlvgkiptelgtcpelflvlsenlltgnlprsfgnlpnlqelqlsvnqlsgtipeelanc
tklthleidnnqlsgeippligklttsltmffawqnqltgtipeslsqcqelqaidlsynnlsgsipngifglefvdlhsngltggIpgtlp
kslqfidlsdnsltgsiptgigsltelfklnlaknrfsgeipreisscrslqllnlgdngftgeipnelgriplaislsInlscnhftgeipsrf
ssltnlgtldvshnklagnlnvladlqnlvslnisfnefsgelpntlIfrklplslvlesnkglfistrpengiqtrhrsavkvtmsilvaa
svvlvlmavytlvkaqritgkqeeldswevtlyqkldfsiddivknltsanvigtgssgvvyrvtipsgetlavkkmwskeenra
fnseintlgsirhrnlrllgwcsnrnlkllfydylpngslssllhgagkgsggadwearydvvlgvahalaylhhdclppilhgdv
kamnvllgsrfesyladfglakivsgegvtdgdsskIsnrpplagsygymapehasmqhiteksdvysygvvllevltgkh
pldpdlpggahlvqwvrdhlagkkdpreildprlrgradpimehmlqtlavsflcvsnkasdrpmmkdivamlkeirqfdm
drsesdmikggkcekwqpqplppekivstprgssncsfaysdesv
```

RCH1.4 AL022223 (T05050 or CAA18216)

```
mppniyrlsffssllcfffipcfsldqqgqallswksqlnisgdafsswhvadtspcnwvgvkcnrrgevseiqlkgmdlqgsl
pvtslrslkslIsltlsslnltgvipkeigdftelelldlsdnslsgdipveifrlkklktlslntnnlegnipmeignlsglvelmlfdnkls
geiprsigelknlqvlraggnknlrgelpweignсnelvmlglaetslsgklpasignlkrvqtialytslIsgpipdeigyctelqn
lylyqnsisgsipttiggllkklqslllwqnnlvgkiptelgncpelwlidfsenlltgtiprsfgklenlqelqlsvnqlsgtipeeltnct
klthleidnnlltgesiplmsnlrsltmffawqnkltgnipqslsqcrelqaidlsynstsgsipkeifglefldlhtnslsgsllgttlpk
slkfidfsdnalsstlppgigllteltklnlaknrlsgeipreistcrslqllnlgendfsgeipdelgqipslaislnlscnrfvgeipsrfs
dlknlgvldvshnqltgnlnvltdlqnlvslnisyndfsgdlpntpfrrlplsdlasnrglyisnaistrpdpttrnssvvrtlillvvvt
avlvlmavytlvraraagkqlIgeeidswevtlyqkldfsiddivknltsanvigtgssgvvyritipsgeslavkkmwskeesg
afnseiktlgsirhrnlvrllgwcsnrnlkllfydylpngslssrlhgagkggcvdwearydvvlgvahalaylhhdclptiihgdv
kamnvllgphfepyladfglartisgypntgidlakptnrppmagslwhgssfdfdlfcllgftehasmqritekscdvysygvv
llevltgkhpldpdlpggahlvkwvrdhlaekkdpsrlldprldgrtdsimhemlqtlavaflcvsnkanerplmkdvvamlte
irhidvgrsetekikaggcgskepqqfmsnekiinshgssncsfafsddsv
```

RCH1.5 AC015446 (AAG12526)

merersnffflflfcswvsmaqptlslssdgqallslkrpspslfsswdpqdqtpcswygitcsadnrvisvsipdtflnlssipdl
sslsslqflnlsstnlsgpippsfgkIthirlldlssnslsgpipselgrlstlqfllInanklsgsipsqisnlfalqvlclqdnllngsips
sfgslvslqqfrlggntnlggpipaqlgflknlttlgfaasglsgsipstfgnlvnlqtlalydteisgtippqlglcselrnlylhmnklt
gsipkelgklqkitslllwgnslsgvippeisncsslvvfdvsandltgdipgdlgklvwleqlqlsdnmftgqipwelsncsslia
lqldknklsgsipsqignlkslqsfflwensisgtipssfgnctdlvaldlsrnkltgripeelfslklsklllignslsgglpksvakcq
slvrlrvgenqlsgqipkeigelqnlvfldlymnhfsgglpyeisnitvlelldvhnnyitgdipaqlgnlvnleqldlsrnsftgnipl
sfgnlsylnklilnnnlitgqipksiknlqkltlldlsynslsgeipqelgqvtsltinldlsyntftgnipetfsdltqlqsldlssnslhgdi
kvlgsltslaslnlscnnfsgpipstpflktisttsylqntnlchsldgitcsshtgqnngvkspkivaltavilasitiailaawlilrnn
hlyktsqnssssspstaedfsypwtfipfqklgitvnnivtsltdenvigkgcsgivykaeipngdivavkklwktkdnneeges
tidsfaaeiqilgnirhrnivkllgycsnksvkllIynyfpngnlqqllqgnrnldwetrykiaigaaqglaylhhdcvpailhrdvk
cnnilldskyealladfglaklmmnspnyhnamsrvaeygytmniteksdvysygvvlleilsgrsavepqigdglhivewv
kkkmgtfepalsvldvklqglpdqivqemlqtlgiamfcvnpspverptmkevvtllmevkcspeewgktsqplikpsss

crop011pct.ST25.txt
SEQUENCE LISTING

<110> UNIVERSITEIT UTRECHT

<120> A novel root specific promoter driving the expression of a novel LRR recepto
r-like
kinase

<130> CROP-011-PCT

<150> EP 00870288.8
<151> 2000-12-04

<150> US 60/256,204
<151> 2000-12-15

<160> 19

<170> PatentIn version 3.1

<210> 1
<211> 7438
<212> DNA
<213> Arabidopsis thaliana

<400> 1
ctcgagtttc agatgtttct attaaataag taatgaaagc agctctaatt ctcccaaaat 60

tcacaagtat agaaagatag aaagatagaa acagagttat tgcatggaag ctgaaaatca 120

tacagcatgg tggatataac caatgagaca caaaacaatc acggcatgtg ttatagaaca 180

acacttatca tctgtattaa cataaagcaa aagttcctat acgacatagt ccaaagttac 240

agaccataag cacaccttat gatttgttgt accctctaaa acacaggatg atgacaagaa 300

gcaaatttgc ccaaatttac tgagggaaaa acgagggctt aacgctctca tatctcaaac 360

aaaaacttct caaacagttt catgtgttca gattggaggg caatggcaac ggataagctg 420

ccatcattag taggacttgg tagcacaaaa gacaaaccct cgtatggaat tccaccaggt 480

cccataaaga taggacgacc ccaaccaaag tctgcatcat aaataggtaa tctaacccag 540

cttgtgattc ccaaatttgg gcacttgtag gtatgtgcac cgcggacaag ggctgacaga 600

tcaggctgca tctccaggta gtcaagagct gacctcagat agttatcatc catacgaacc 660

aagaaatcat gaatctgtcc tgcagcatac catgttggct tagataacaa atcccctgca 720

acagccaatg gtgttgcagt gaatatcaca ttcccaaagt aaccaggagg cagctgcgga 780

cgtagtctag accttccatc agttgcaatg tacagtttcg tctcttggtc gtttggaagc 840

cctcgcgcct ttcccactga tctccacaca tgccctgcca acatctcgta tgagctgtag 900

ctgacagtgt tcccatcctc cttggatttc gccttaagag caacaagctg gtctcgtgtt 960

aatttgaata tagagacagt ggtattctca ggtcctgatt tagacggatc aagaggtatc 1020

ttcatacttg gtgcaggctg atattcaaca tgatgaaaag caggctgtgg cgggtcccta 1080

gctcggagga gtgttcgatc aatgaaaggt ggaatggtta ggtcaagacc acgagccata 1140

tcagaccatg tgttgataaa atgaagacca gagaaaccat ctgccgcgtg atgttgcatc 1200
ccaacccccaa gtgaagctcc cccacattta aagaaagtca cctatatcat caaacaatat 1260
cattacataa ctaaaaacag aacatacatc atcaaagaat atcattacag aactaaaaaa 1320
cagaacaaca caagcattca ataaattaaa aaagggaaag actttaaatt aagctgacct 1380
gcaaaacgag aagcgggaaa gagtgaatgc cagcggagtg atcaacttcg ggaataagct 1440
gacggagatt aaggggtagga gcaaaatcac caaaatcatc gataacagaa ggagtatcag 1500
ccacaacgaa gagaacacca gcaccgttac aatcgatctc aatacgacca tcatcgtctc 1560
tcttcaagcg accagccata gggtaaaaag ggacaagggc tttggaaaga gcttccttca 1620
tgacctgagg gtcaaagaaa ttggaagcgc cggtgggtct gtagaagtag acactagggg 1680
tatggaatct ggggatgaca aggtcgacgt tggagttcca aagattagtg attggtgtct 1740
cggtggcagg ccggaccatg gtggaatctc tgatgttaat tttcatattt ttaagaaaga 1800
tcaaacgaac aaacaaattt tgttttgagt ttgatctgag aaattgtgtg agaagcgagg 1860
aggggtagat ataagtgggt gtgaacaaaa aaggaaaggt cgtaatcctt cacctaccgc 1920
cccacgaccc cgcctactcc atctgtcgac caactctctc agaaggttta tcattggtcc 1980
atctcatcac tcggtcctac ttttctatcc ggggttggta aacccgttcc ctctgtccta 2040
aggcccaaat aatactaatg ggcttaaata gtataaagtt cgaatgaaga agaaggaaaa 2100
aaaattgaca agttgattgt tatagctacc gtacaagtta gtcaatcgaa tgccaaagtt 2160
tggcttaact taactcatta ccacgtagca gagcataaca acaaccacca ataacatcta 2220
acggcgtaaa ctcacgtgtt aagagaatta cgtcagaatc caagaaagtc tttttttcag 2280
tacaccacaa ataaattatg ttgaatttat atccgaacta tactatatta tatgaaaaca 2340
tacctcaatt agtcataaat ttcacatttc attttaattt attaactaat ggttgttgtt 2400
gcggatagct tttctttgct ttctgttcta aaattatttc ttagccatta tttgatgatc 2460
caacttggtt tgaggatcta ataaaaattc aatttctttg aagcattcat tggcttaatg 2520
atccaaacct tttgtcctcc cactctctaa gaaaataaag caaaagcaaa actaaaagaa 2580
aagcaactag taccgtgggt tggtcaatgg taacccatgc taccaaatac caatcaaagc 2640
ttaaatgaca atttagtacc ttgggttggt catgatttag agcggaacaa atataccata 2700
catcaaacga ggatatacag agaaaattca tggaagtatg gaatttagag gacaatttct 2760
cttctgggct acaacggacc ggcccattcg ctcatttacc cagaggtatc gagtttgtgg 2820
acttttgatg ccgctagaga ctattggcat cggattgaaa aaaatgttta cttcgttgtt 2880
aacaattttc tgaatgcaat attttccttg tcatgaatat ttaaacttgt tattactttc 2940
ttttagctta ggtgtggaca attatggagt ttacttcaaa cgaggaagaa tcttaaacgc 3000

tcggttcagg tctcgaaaac aaaccaactc acaatcctga cttaattgag gaaaacaatg 3060

caaaaccaca tgcatgcttc catatttcta tcataatctt ataagaaaaa acactactaa 3120

gtgaaatgat tctgtatata tataaccaat gccttttgtt ttgtgatatt ttatgtatat 3180

ataactattg acttttgtca tctatggata gtgtctcggg ctcttggcaa acatatttca 3240

aagaaaagtt aatgactgta attaattaat ctgaagctag aaacagaacc ccgaggtaaa 3300

agaaaaagac agagcacatg aagtttagta cttttatata tttaatatat cattctttct 3360

tattgcttat ctctaaagca aaaacttccc taaaccctaa gccaaaggac tcagatcgat 3420

gcagaaccaa gaaggcttgt tttggatttg agagccaaat gcaaagaaaa aaactcttat 3480

gtctttgcaa atgccgattc caagaaaaaa agccttaact gtttctcatt tttccattac 3540

attgtctctg tttttagctt tcttcatctc ctcgacctct gcatcaacca atgaagtctc 3600

agctttgatt tcttggcttc acagctcgaa ctcgccgcca ccgtcagtct tctccggctg 3660

gaatccttct gattctgacc catgtcaatg gccttacatt acttgttcct cctcagacaa 3720

caaactcgtt acagagatca atgtcgtctc tgttcagtta gctttacctt tccctcctaa 3780

catttcttcg tttacttcac ttcagaaact cgttatctcc aacactaatc tcaccggagc 3840

tatatcttct gagatcggag attgttcgga gcttatagtt atcgatttaa gctcaaatag 3900

tcttgttggt gaaatacctt cgagtctagg gaagctcaag aatcttcaag aactttgttt 3960

aaactccaat ggcctcacag gaaagatccc accggaactc ggcgactgcg ttagcctcaa 4020

gaatcttgag attttcgata actacttatc cgagaatctc ccgttggagc tcggaaagat 4080

ctcgactctt gagagcataa gagccggagg aaactcagag ctttcaggga agatcccgga 4140

ggagatcgga aactgtagga atctcaaggt cttagggtta gcagcaacga aaatatccgg 4200

ttctttacct gtttcgttgg gtcaactaag caagctccag tctttatctg tctattctac 4260

aatgctctca ggtgagatcc ctaaagagct tggaaactgc tctgaactta tcaatctgtt 4320

tctatacgac aatgacttgt ccggcacgct tccaaaagaa ctaggaaagc ttcaaaacct 4380

agagaagatg cttctatggc agaataatct ccacggacct atccctgagg agattggatt 4440

catgaaaagc ttaaacgcca ttgatctctc tatgaactat ttctcaggaa ccatccctaa 4500

atcgttcggt aacttgtcga atctccaaga gcttatgctt agcagcaaca acatcaccgg 4560

gtcaatacct tcgattctaa gtaactgcac aaagcttgtt cagttccaga ttgacgccaa 4620

tcagatttca ggtttgattc caccagagat tggattgctt aaggagctca acatcttctt 4680

gggatggcag aataagctag aagggaatat tccggacgag ttagctggtt gtcagaatct 4740

tcaagctctt gacttgtcac agaattatct caccggatca ttacctgcag gtttgtttca 4800

gcttcgtaat ctgactaagc tcttgcttat atctaatgcc atctccggtg ttatcccact 4860

ggagatcggg aactgcactt cgcttgtcag attaaggctt gtcaataaca gaatcaccgg 4920

agaaatccct aaaggtattg gatttcttca gaaccttagc ttcctcgact tgtctgagaa 4980
caatctctcc ggtccagttc ctttggagat aagtaattgt agacaactcc agatgcttaa 5040
cttgagcaac aacactcttc aaggttatct tcctctgtct ttatcatcgt taacaaagct 5100
tcaagtactt gatgtcttct ccaatgactt gactggcaag atacctgata gtctaggtca 5160
tctcatttca ctcaaccggc tcattctaag taagaattct ttcaatggag aaatcccttc 5220
gtctctcggt cactgcacga atcttcagct tcttgatctc agcagtaata acatctctgg 5280
aactataccg gaggagcttt ttgacatcca agatctagac attgcgttaa acttgagctg 5340
gaattcatta gatggcttta tcccggaaag gatttccgcg cttaaccgat tatccgtgct 5400
cgatatttcg cataacatgc tttcaggcga cctttctgcg ctgtccggtc tagaaaactt 5460
ggtttctctg aacatctctc ataacagatt ctcaggttat cttccagata gtaaagtgtt 5520
tagacagctg ataggagcag agatggaagg aaacaatgga ctctgttcca aaggtttcag 5580
gtcttgcttt gtaagtaaca gttcacagtt aactacacag cgtggtgtgc actcacatag 5640
actcaggata gccattggat tgctaatcag cgtgacagcg gttctagcgg tattaggcgt 5700
gttggcggtt atacgagcta aacaaatgat tcgagatgat aacgattcag agactggaga 5760
aaatctatgg acatggcaat tcacaccttt tcagaaactc aacttcacag tcgaacatgt 5820
actcaagtgt ttggtagaag gtaatgttat agggaaaggt tgctccggga tagtgtacaa 5880
agctgaaatg cctaacagag aagtcatcgc ggtgaaaaag ctctggccag tgacagtacc 5940
taatctgaat gagaaaacta agtcatcagg agttcgagac tcattctcag ctgaagtaaa 6000
aacacttgga tcgatcagac acaagaacat tgtacgcttc ttgggatgtt gttggaacaa 6060
gaacactaga cttcttatgt atgattatat gtcaaatggg agtttgggaa gtttgcttca 6120
cgaaaggagc ggtgtatgta gcttaggatg ggaagtgagg tacaagatta tacttggtgc 6180
agctcagggt ttggcttact tgcaccatga ctgtgttcct cccattgttc atagagacat 6240
caaggcaaat aatattctga ttggccctga ttttgaacct tacattggag atttcggact 6300
cgctaagctt gttgatgatg gcgactttgc tcgttcttcc aacaccattg ctggttccta 6360
tggttacata gctccaggta cttcccaaat ttctgttttg ctgctaatta gttaacttga 6420
tacatcataa gttactagta gtcttagcct agtaatggta gaaattcagg tttatagttt 6480
ggttttggtt tttggcattt tgaaaagaat agtattttgg ttcagttgat tttcgatttg 6540
gtttaaaatt ttagaaaaaa aatggtagaa gaaataattt ttgattaata ttatatcatc 6600
tactgtatac gatttaaaaa aaaaattgat ttggtttggt ctcgatattt tgcttgtagc 6660
ttggttcaat atttttacta tatatatgat tcatgacttt catactaatc aatatttgtt 6720
cggtttcgtt ttgattagtc tgaaagtgaa tcatatagtt aaaatattga ccaatctatc 6780

atatacttac acgctacaag attttcgatg taattttggt gtaagcattt tggtttttat 6840

atgttcatcc ttagtaacta acaacaaata tcaattggga agcataacgt tatatattcc 6900

cataaggcta taacactata attatgcaga atacggatac tcaatgaaga taacagagaa 6960

aagcgacgtg tacagctatg gagtcgtagt gctcgaggta ctaacgggta agcaaccaat 7020

cgatccaacg ataccagatg gactccacat agtggactgg gtcaagaaaa tcagagacat 7080

acaagtaatc gaccaaggac tacaagcaag accagagtca gaggttgaag agatgatgca 7140

aacgctagga gtcgcgcttt tatgcattaa tccaataccc gaagacaggc ctacaatgaa 7200

agatgtggct gctatgctta gtgagatatg tcaagagaga gaggaatcga tgaaagttga 7260

tggttgctcg ggaagttgta acaatggaag agaacgtggc aaagatgatt cgacttcatc 7320

ggttatgcaa caaacggcta agtatttgag aagtagtagc acgagtttct ctgcgtcttc 7380

tttgctttac tcttcttctt cttctgctac ttctaatgtt agaccaaatc ttaaataa 7438

<210> 2
<211> 3499
<212> DNA
<213> Arabidopsis thaliana

<400> 2
ctcgagtttc agatgtttct attaaataag taatgaaagc agctctaatt ctcccaaaat 60

tcacaagtat agaaagatag aaagatagaa acagagttat tgcatggaag ctgaaaatca 120

tacagcatgg tggatataac caatgagaca caaaacaatc acggcatgtg ttatagaaca 180

acacttatca tctgtattaa cataaagcaa aagttcctat acgacatagt ccaaagttac 240

agaccataag cacaccttat gatttgttgt accctctaaa acacaggatg atgacaagaa 300

gcaaatttgc ccaaatttac tgagggaaaa acgagggctt aacgctctca tatctcaaac 360

aaaaacttct caaacagttt catgtgttca gattggaggg caatggcaac ggataagctg 420

ccatcattag taggacttgg tagcacaaaa gacaaaccct cgtatggaat tccaccaggt 480

cccataaaga taggacgacc ccaaccaaag tctgcatcat aaataggtaa tctaacccag 540

cttgtgattc ccaaatttgg gcacttgtag gtatgtgcac cgcggacaag ggctgacaga 600

tcaggctgca tctccaggta gtcaagagct gacctcagat agttatcatc catacgaacc 660

aagaaatcat gaatctgtcc tgcagcatac catgttggct tagataacaa atcccctgca 720

acagccaatg gtgttgcagt gaatatcaca ttcccaaagt aaccaggagg cagctgcgga 780

cgtagtctag accttccatc agttgcaatg tacagtttcg tctcttggtc gtttggaagc 840

cctcgcgcct ttcccactga tctccacaca tgccctgcca acatctcgta tgagctgtag 900

ctgacagtgt tcccatcctc cttggatttc gccttaagag caacaagctg gtctcgtgtt 960

aatttgaata tagagacagt ggtattctca ggtcctgatt tagacggatc aagaggtatc 1020

ttcatacttg gtgcaggctg atattcaaca tgatgaaaag caggctgtgg cgggtcccta 1080

gctcggagga gtgttcgatc aatgaaaggt ggaatggtta ggtcaagacc acgagccata 1140

tcagaccatg tgttgataaa atgaagacca gagaaaccat ctgccgcgtg atgttgcatc 1200

ccaaccccaa gtgaagctcc cccacattta aagaaagtca cctatatcat caaacaatat 1260

cattacataa ctaaaaacag aacatacatc atcaaagaat atcattacag aactaaaaaa 1320

cagaacaaca caagcattca ataaattaaa aaagggaaag actttaaatt aagctgacct 1380

gcaaaacgag aagcgggaaa gagtgaatgc cagcggagtg atcaacttcg ggaataagct 1440

gacggagatt aagggtagga gcaaaatcac caaaatcatc gataacagaa ggagtatcag 1500

ccacaacgaa gagaacacca gcaccgttac aatcgatctc aatacgacca tcatcgtctc 1560

tcttcaagcg accagccata gggtaaaaag ggacaagggc tttggaaaga gcttccttca 1620

tgacctgagg gtcaaagaaa ttggaagcgc cggtgggtct gtagaagtag acactagggg 1680

tatggaatct ggggatgaca aggtcgacgt tggagttcca aagattagtg attggtgtct 1740

cggtggcagg ccggaccatg gtggaatctc tgatgttaat tttcatattt ttaagaaaga 1800

tcaaacgaac aaacaaattt tgttttgagt ttgatctgag aaattgtgtg agaagcgagg 1860

aggggtagat ataagtgggt gtgaacaaaa aaggaaaggt cgtaatcctt cacctaccgc 1920

cccacgaccc cgcctactcc atctgtcgac caactctctc agaaggttta tcattggtcc 1980

atctcatcac tcggtcctac ttttctatcc ggggttggta aacccgttcc ctctgtccta 2040

aggcccaaat aatactaatg ggcttaaata gtataaagtt cgaatgaaga agaaggaaaa 2100

aaaattgaca agttgattgt tatagctacc gtacaagtta gtcaatcgaa tgccaaagtt 2160

tggcttaact taactcatta ccacgtagca gagcataaca acaaccacca ataacatcta 2220

acggcgtaaa ctcacgtgtt aagagaatta cgtcagaatc caagaaagtc ttttttcag 2280

tacaccacaa ataaattatg ttgaatttat atccgaacta tactatatta tatgaaaaca 2340

tacctcaatt agtcataaat ttcacatttc attttaattt attaactaat ggttgttgtt 2400

gcggatagct tttctttgct ttctgttcta aaattatttc ttagccatta tttgatgatc 2460

caacttggtt tgaggatcta ataaaaattc aatttctttg aagcattcat tggcttaatg 2520

atccaaacct tttgtcctcc cactctctaa gaaaataaag caaaagcaaa actaaaagaa 2580

aagcaactag taccgtgggt tggtcaatgg taacccatgc taccaaatac caatcaaagc 2640

ttaaatgaca atttagtacc ttgggttggt catgatttag agcggaacaa atataccata 2700

catcaaacga ggatatacag agaaaattca tggaagtatg gaatttagag gacaatttct 2760

cttctgggct acaacggacc ggcccattcg ctcatttacc cagaggtatc gagtttgtgg 2820

acttttgatg ccgctagaga ctattggcat cggattgaaa aaaatgttta cttcgttgtt 2880

aacaattttc tgaatgcaat attttccttg tcatgaatat ttaaacttgt tattactttc 2940

tttta[illegible] ggtgtggaca attatggagt ttacttcaaa cgaggaagaa tcttaaacgc 3000

tcggt[illegible] tctcgaaaac aaaccaactc acaatcctga cttaattgag gaaaacaatg 3060

caaaac[illegible] tgcatgcttc catatttcta tcataatctt ataagaaaaa acactactaa 3120

gtqaa[illegible] tctgtatata tataaccaat gcctttttgtt ttgtgatatt ttatgtatat 3180

ataa[illegible] actttttgtca tctatggata gtgtctcggg ctcttggcaa acatatttca 3240

aaqa[illegible] aatgactgta attaattaat ctgaagctag aaacagaacc ccgaggtaaa 3300

ags[illegible] agagcacatg aagtttagta cttttatata tttaatatat cattctttct 3360

tatt[illegible] ctctaaagca aaaacttccc taaaccctaa gccaaaggac tcagatcgat 3420

gca[illegible] gaaggcttgt tttggatttg agagccaaat gcaaagaaaa aaactcttat 3480

gtc[illegible] atgccgatt 3499

<210> 1
<211> 3560
<212> DNA
<213> Arabidopsis thaliana

<400> 1
atg[illegible] aaatgccgat tccaagaaaa aaagccttaa ctgtttctca tttttccatt 60

acatt[illegible] tgtttttagc tttcttcatc tcctcgacct ctgcatcaac caatgaagtc 120

tca[illegible] tttcttggct tcacagctcg aactcgccgc caccgtcagt cttctccggc 180

tg[illegible] ctgattctga cccatgtcaa tggccttaca ttacttgttc ctcctcagac 240

aa[illegible] ttacagagat caatgtcgtc tctgttcagt tagctttacc tttccctcct 300

aaca[illegible] cgtttacttc acttcagaaa ctcgttatct ccaacactaa tctcaccgga 360

gct[illegible] ctgagatcgg agattgttcg gagcttatag ttatcgattt aagctcaaat 420

ag[illegible] gtgaaatacc ttcgagtcta gggaagctca agaatcttca agaactttgt 480

ttaaa[illegible] atggcctcac aggaaagatc ccaccggaac tcggcgactg cgttagcctc 540

aaja[illegible] agattttcga taactactta tccgagaatc tcccgttgga gctcggaaag 600

at[illegible] ttgagagcat aagagccgga ggaaactcag agctttcagg gaagatcccg 660

gt[illegible] gaaactgtag gaatctcaag gtcttagggt tagcagcaac gaaaatatcc 720

gg[illegible] ctgtttcgtt gggtcaacta agcaagctcc agtctttatc tgtctattct 780

acta[illegible] caggtgagat ccctaaagag cttggaaact gctctgaact tatcaatctg 840

ttt[illegible] acaatgactt gtccggcacg cttccaaaag aactaggaaa gcttcaaaac 900

cta[illegible] tgcttctatg gcagaataat ctccacggac ctatccctga ggagattgga 960

ttca[illegible] gcttaaacgc cattgatctc tctatgaact atttctcagg aaccatccct 1020

aaa[illegible] gtaacttgtc gaatctccaa gagcttatgc ttagcagcaa caacatcacc 1080

gggtcaatac cttcgattct aagtaactgc acaaagcttg ttcagttcca gattgacgcc 1140
aatcagattt caggtttgat tccaccagag attggattgc ttaaggagct caacatcttc 1200
ttgggatggc agaataagct agaagggaat attccggacg agttagctgg ttgtcagaat 1260
cttcaagctc ttgacttgtc acagaattat ctcaccggat cattacctgc aggtttgttt 1320
cagcttcgta atctgactaa gctcttgctt atatctaatg ccatctccgg tgttatccca 1380
ctggagatcg ggaactgcac ttcgcttgtc agattaaggc ttgtcaataa cagaatcacc 1440
ggagaaatcc ctaaaggtat tggatttctt cagaacctta gcttcctcga cttgtctgag 1500
aacaatctct ccggtccagt tcctttggag ataagtaatt gtagacaact ccagatgctt 1560
aacttgagca acaacactct tcaaggttat cttcctctgt ctttatcatc gttaacaaag 1620
cttcaagtac ttgatgtctc ttccaatgac ttgactggca agatacctga tagtctaggt 1680
catctcattt cactcaaccg gctcattcta agtaagaatt ctttcaatgg agaaatccct 1740
tcgtctctcg gtcactgcac gaatcttcag cttcttgatc tcagcagtaa taacatctct 1800
ggaactatac cggaggagct ttttgacatc caagatctag acattgcgtt aaacttgagc 1860
tggaattcat tagatggctt tatcccggaa aggatttccg cgcttaaccg attatccgtg 1920
ctcgatattt cgcataacat gctttcaggc gacctttctg cgctgtccgg tctagaaaac 1980
ttggtttctc tgaacatctc tcataacaga ttctcaggtt atcttccaga tagtaaagtg 2040
tttagacagc tgataggagc agagatggaa ggaaacaatg gactctgttc caaaggtttc 2100
aggtcttgct ttgtaagtaa cagttcacag ttaactacac agcgtggtgt gcactcacat 2160
agactcagga tagccattgg attgctaatc agcgtgacag cggttctagc ggtattaggc 2220
gtgttggcgg ttatacgagc taaacaaatg attcgagatg ataacgattc agagactgga 2280
gaaaatctat ggacatggca attcacacct tttcagaaac tcaacttcac agtcgaacat 2340
gtactcaagt gtttggtaga aggtaatgtt atagggaaag gttgctccgg gatagtgtac 2400
aaagctgaaa tgcctaacag agaagtcatc gcggtgaaaa agctctggcc agtgacagta 2460
cctaatctga atgagaaaac taagtcatca ggagttcgag actcattctc agctgaagta 2520
aaaacacttg gatcgatcag acacaagaac attgtacgct tcttgggatg ttgttggaac 2580
aagaacacta gacttcttat gtatgattat atgtcaaatg ggagtttggg aagtttgctt 2640
cacgaaagga gcggtgtatg tagcttagga tgggaagtga ggtacaagat tatacttggt 2700
gcagctcagg gtttggctta cttgcaccat gactgtgttc ctcccattgt tcatagagac 2760
atcaaggcaa ataatattct gattggccct gattttgaac cttacattgg agatttcgga 2820
ctcgctaagc ttgttgatga tggcgacttt gctcgttctt ccaacaccat tgctggttcc 2880
tatggttaca tagctccagg tacttcccaa atttctgttt tgctgctaat tagttaactt 2940

gatacatcat aagttactag tagtcttagc ctagtaatgg tagaaattca ggtttatagt 3000

ttggttttgg tttttggcat tttgaaaaga atagtatttt ggttcagttg attttcgatt 3060

tggtttaaaa ttttagaaaa aaaatggtag aagaaataat ttttgattaa tattatatca 3120

tctactgtat acgatttaaa aaaaaaattg atttggtttg gtctcgatat tttgcttgta 3180

gcttggttca atatttttac tatatatatg attcatgact ttcatactaa tcaatatttg 3240

ttcggtttcg ttttgattag tctgaaagtg aatcatatag ttaaaatatt gaccaatcta 3300

tcatatactt acacgctaca agattttcga tgtaattttg gtgtaagcat tttggttttt 3360

atatgttcat ccttagtaac taacaacaaa tatcaattgg gaagcataac gttatatatt 3420

cccataaggc tataacacta taattatgca gaatacggat actcaatgaa gataacagag 3480

aaaagcgacg tgtacagcta tggagtcgta gtgctcgagg tactaacggg taagcaacca 3540

atcgatccaa cgataccaga tggactccac atagtggact gggtcaagaa aatcagagac 3600

atacaagtaa tcgaccaagg actacaagca agaccagagt cagaggttga agagatgatg 3660

caaacgctag gagtcgcgct tttatgcatt aatccaatac ccgaagacag gcctacaatg 3720

aaagatgtgg ctgctatgct tagtgagata tgtcaagaga gagaggaatc gatgaaagtt 3780

gatggttgct cgggaagttg taacaatgga agagaacgtg gcaaagatga ttcgacttca 3840

tcggttatgc aacaaacggc taagtatttg agaagtagta gcacgagttt ctctgcgtct 3900

tctttgcttt actcttcttc ttcttctgct acttctaatg ttagaccaaa tcttaaataa 3960

<210> 4
<211> 1135
<212> PRT
<213> Arabidopsis thaliana

<400> 4

Met Ser Leu Gln Met Pro Ile Pro Arg Lys Lys Ala Leu Thr Val Ser
1 5 10 15

His Phe Ser Ile Thr Leu Ser Leu Phe Leu Ala Phe Phe Ile Ser Ser
20 25 30

Thr Ser Ala Ser Thr Asn Glu Val Ser Ala Leu Ile Ser Trp Leu His
35 40 45

Ser Ser Asn Ser Pro Pro Pro Ser Val Phe Ser Gly Trp Asn Pro Ser
50 55 60

Asp Ser Asp Pro Cys Gln Trp Pro Tyr Ile Thr Cys Ser Ser Pro Asp
65 70 75 80

Asn Lys Leu Val Thr Glu Ile Asn Val Val Ser Val Gln Leu Ala Leu

85 90 95

Pro Phe Pro Pro Asn Ile Ser Ser Phe Thr Ser Leu Gln Lys Leu Val
100 105 110

Ile Ser Asn Thr Asn Leu Thr Gly Ala Ile Ser Ser Glu Ile Gly Asp
115 120 125

Cys Ser Glu Leu Ile Val Ile Asp Leu Ser Ser Asn Ser Leu Val Gly
130 135 140

Glu Ile Pro Ser Ser Leu Gly Lys Leu Lys Asn Leu Gln Glu Leu Cys
145 150 155 160

Leu Asn Ser Asn Gly Leu Thr Gly Lys Ile Pro Pro Glu Leu Gly Asp
165 170 175

Cys Val Ser Leu Lys Asn Leu Glu Ile Phe Asp Asn Tyr Leu Ser Glu
180 185 190

Asn Leu Pro Leu Glu Leu Gly Lys Ile Ser Thr Leu Glu Ser Ile Arg
195 200 205

Ala Gly Gly Asn Ser Glu Leu Ser Gly Lys Ile Pro Glu Glu Ile Gly
210 215 220

Asn Cys Arg Asn Leu Lys Val Leu Gly Leu Ala Ala Thr Lys Ile Ser
225 230 235 240

Gly Ser Leu Pro Val Ser Leu Gly Gln Leu Ser Lys Leu Gln Ser Leu
245 250 255

Phe Val Tyr Ser Thr Met Leu Ser Gly Glu Ile Pro Lys Glu Leu Gly
260 265 270

Asn Cys Ser Glu Leu Ile Asn Leu Phe Leu Tyr Asp Asn Asp Leu Ser
275 280 285

Gly Thr Leu Pro Lys Glu Leu Gly Lys Leu Gln Asn Leu Glu Lys Met
290 295 300

Leu Leu Trp Gln Asn Asn Leu His Gly Pro Ile Pro Glu Glu Ile Gly
305 310 315 320

Phe Met Lys Ser Leu Asn Ala Ile Asp Leu Ser Met Asn Tyr Phe Ser
325 330 335

Gly Thr Ile Pro Lys Ser Phe Gly Asn Leu Ser Asn Leu Gln Glu Leu
340 345 350

Met Leu Ser Ser Asn Asn Ile Thr Gly Ser Ile Pro Ser Ile Leu Ser
355 360 365

Asp Cys Thr Lys Leu Val Gln Phe Gln Ile Asp Ala Asn Gln Ile Ser
370 375 380

Gly Leu Ile Pro Pro Glu Ile Gly Leu Leu Lys Glu Leu Asn Ile Phe
385 390 395 400

Leu Gly Trp Gln Asn Lys Leu Glu Gly Asn Ile Pro Asp Glu Leu Ala
405 410 415

Gly Cys Gln Asn Leu Gln Ala Leu Asp Leu Ser Gln Asn Tyr Leu Thr
420 425 430

Gly Ser Leu Pro Ala Gly Leu Phe Gln Leu Arg Asn Leu Thr Lys Leu
435 440 445

Leu Leu Ile Ser Asn Ala Ile Ser Gly Val Ile Pro Leu Glu Thr Gly
450 455 460

Asn Cys Thr Ser Leu Val Arg Leu Arg Leu Val Asn Asn Arg Ile Thr
465 470 475 480

Gly Glu Ile Pro Lys Gly Ile Gly Phe Leu Gln Asn Leu Ser Phe Leu
485 490 495

Asp Leu Ser Glu Asn Asn Leu Ser Gly Pro Val Pro Leu Glu Ile Ser
500 505 510

Asn Cys Arg Gln Leu Gln Met Leu Asn Leu Ser Asn Asn Thr Leu Gln
515 520 525

Gly Tyr Leu Pro Leu Ser Leu Ser Ser Leu Thr Lys Leu Gln Val Leu
530 535 540

Asp Val Ser Ser Asn Asp Leu Thr Gly Lys Ile Pro Asp Ser Leu Gly
545 550 555 560

His Leu Ile Ser Leu Asn Arg Leu Ile Leu Ser Lys Asn Ser Phe Asn
565 570 575

Gly Glu Ile Pro Ser Ser Leu Gly His Cys Thr Asn Leu Gln Leu Leu
580 585 590

Asp Leu Ser Ser Asn Asn Ile Ser Gly Thr Ile Pro Glu Glu Leu Phe
595 600 605

Asp Ile Gln Asp Leu Asp Ile Ala Leu Asn Leu Ser Trp Asn Ser Leu
610 615 620

Asp Gly Phe Ile Pro Glu Arg Ile Ser Ala Leu Asn Arg Leu Ser Val
625 630 635 640

Leu Asp Ile Ser His Asn Met Leu Ser Gly Asp Leu Ser Ala Leu Ser
645 650 655

Gly Leu Glu Asn Leu Val Ser Leu Asn Ile Ser His Asn Arg Phe Ser
660 665 670

Gly Tyr Leu Pro Asp Ser Lys Val Phe Arg Gln Leu Ile Gly Ala Glu
675 680 685

Met Glu Gly Asn Asn Gly Leu Cys Ser Lys Gly Phe Arg Ser Cys Phe
690 695 700

Val Ser Asn Ser Ser Gln Leu Thr Thr Gln Arg Gly Val His Ser His
705 710 715 720

Arg Leu Arg Ile Ala Ile Gly Leu Leu Ile Ser Val Thr Ala Val Leu
725 730 735

Ala Val Leu Gly Val Leu Ala Val Ile Arg Ala Lys Gln Met Ile Arg
740 745 750

Asp Asp Asn Asp Ser Glu Thr Gly Glu Asn Leu Trp Thr Trp Gln Phe
755 760 765

Thr Pro Phe Gln Lys Leu Asn Phe Thr Val Glu His Val Leu Lys Cys
770 775 780

Leu Val Glu Gly Asn Val Ile Gly Lys Gly Cys Ser Gly Ile Val Tyr
785 790 795 800

Lys Ala Glu Met Pro Asn Arg Glu Val Ile Ala Val Lys Lys Leu Trp
805 810 815

Pro Val Thr Val Pro Asn Leu Asn Glu Lys Thr Lys Ser Ser Gly Val
820 825 830

Arg Asp Ser Phe Ser Ala Glu Val Lys Thr Leu Gly Ser Ile Arg His
835 840 845

```
Lys Asn Ile Val Arg Phe Leu Gly Cys Cys Trp Asn Lys Asn Thr Arg
  850                 855                 860

Leu Leu Met Tyr Asp Tyr Met Ser Asn Gly Ser Leu Gly Ser Leu Leu
865                 870                 875                 880

His Glu Arg Ser Gly Val Cys Ser Leu Gly Trp Glu Val Arg Tyr Lys
                885                 890                 895

Ile Ile Leu Gly Ala Ala Gln Gly Leu Ala Tyr Leu His His Asp Cys
            900                 905                 910

Val Pro Pro Ile Val His Arg Asp Ile Lys Ala Asn Asn Ile Leu Ile
        915                 920                 925

Gly Pro Asp Phe Glu Pro Tyr Ile Gly Asp Phe Gly Leu Ala Lys Leu
    930                 935                 940

Val Asp Asp Gly Asp Phe Ala Arg Ser Ser Asn Thr Ile Ala Gly Ser
945                 950                 955                 960

Tyr Gly Tyr Ile Ala Pro Glu Tyr Gly Tyr Ser Met Lys Ile Thr Glu
                965                 970                 975

Lys Ser Asp Val Tyr Ser Tyr Gly Val Val Val Leu Glu Val Leu Thr
            980                 985                 990

Gly Lys Gln Pro Ile Asp Pro Thr  Ile Pro Asp Gly Leu  His Ile Val
        995                 1000                1005

Asp Trp  Val Lys Lys Ile Arg  Asp Ile Gln Val Ile  Asp Gln Gly
    1010                1015                1020

Leu Gln  Ala Arg Pro Glu Ser  Glu Val Glu Glu Met  Met Gln Thr
    1025                1030                1035

Leu Gly  Val Ala Leu Leu Cys  Ile Asn Pro Ile Pro  Glu Asp Arg
    1040                1045                1050

Pro Thr  Met Lys Asp Val Ala  Ala Met Leu Ser Glu  Ile Cys Gln
    1055                1060                1065

Glu Arg  Glu Glu Ser Met Lys  Val Asp Gly Cys Ser  Gly Ser Cys
    1070                1075                1080

Asn Asn  Gly Arg Glu Arg Gly  Lys Asp Asp Ser Thr  Ser Ser Val
```

1085 1090 1095

Met Gln Gln Thr Ala Lys Tyr Leu Arg Ser Ser Ser Thr Ser Phe
1100 1105 1110

Ser Ala Ser Ser Leu Leu Tyr Ser Ser Ser Ser Ser Ala Thr Ser
1115 1120 1125

Asn Val Arg Pro Asn Leu Lys
1130 1135

<210> 5
<211> 25
<212> PRT
<213> Arabidopsis thaliana

<400> 5

Arg Tyr Lys Ile Ile Leu Gly Ala Ala Gln Gly Leu Ala Tyr Leu His
1 5 10 15

His Asp Cys Val Pro Pro Ile Val His
20 25

<210> 6
<211> 23
<212> DNA
<213> Arabidopsis thaliana

<400> 6
aayaayttma syggtkmdwt hcc 23

<210> 7
<211> 20
<212> DNA
<213> Arabidopsis thaliana

<400> 7
carythamcg gwgaaathcc 20

<210> 8
<211> 20
<212> DNA
<213> Arabidopsis thaliana

<400> 8
mayytsavyg gahctathcc 20

<210> 9
<211> 22
<212> DNA
<213> Arabidopsis thaliana

<220>

<221> misc_feature
<222> (17)..(17)
<223> n = a, c, g or t

<400> 9
thccwttwgs catgtantca ta 22

<210> 10
<211> 19
<212> DNA
<213> Arabidopsis thaliana

<400> 10
aytcdggdgc hahgtadcc 19

<210> 11
<211> 21
<212> DNA
<213> Arabidopsis thaliana

<400> 11
mayhccraar ctrtavacat c 21

<210> 12
<211> 19
<212> DNA
<213> Arabidopsis thaliana

<400> 12
cgatcagaca caagaacat 19

<210> 13
<211> 19
<212> DNA
<213> Arabidopsis thaliana

<400> 13
agcaatggtg ttggaagaa 19

<210> 14
<211> 19
<212> DNA
<213> Arabidopsis thaliana

<400> 14
tgcagatctt ygtgaagac 19

<210> 15
<211> 18
<212> DNA
<213> Arabidopsis thaliana

<400> 15
gactccttct ggatgttg 18

<210> 16
<211> 18
<212> DNA
<213> Arabidopsis thaliana

<400> 16
tgtaaaacga cggccagt 18

<210> 17
<211> 34
<212> DNA
<213> Arabidopsis thaliana

<400> 17
cacacaggat ccaatcggca tttgcaaaga cata 34

<210> 18
<211> 842
<212> DNA
<213> Arabidopsis thaliana

<400> 18
aagcttaaat gacaatttag taccttgggt tggtcatgat ttagagcgga acaaatatac 60

catacatcaa acgaggatat acagagaaaa ttcatggaag tatggaattt agaggacaat 120

ttctcttctg ggctacaacg gaccggccca ttcgctcatt tacccagagg tatcgagttt 180

gtggactttt gatgccgcta gagactattg gcatcggatt gaaaaaaatg tttacttcgt 240

tgttaacaat tttctgaatg caatattttc cttgtcatga atatttaaac ttgttattac 300

tttcttttag cttaggtgtg gacaattatg gagtttactt caaacgagga agaatcttaa 360

acgctcggtt caggtctcga aaacaaacca actcacaatc ctgacttaat tgaggaaaac 420

aatgcaaaac cacatgcatg cttccatatt tctatcataa tcttataaga aaaaacacta 480

ctaagtgaaa tgattctgta tatatataac caatgccttt tgttttgtga tattttatgt 540

atatataact attgactttt gtcatctatg gatagtgtct cgggctcttg gcaaacatat 600

ttcaaagaaa agttaatgac tgtaattaat taatctgaag ctagaaacag aaccccgagg 660

taaaagaaaa agacagagca catgaagttt agtacttttа tatatttaat atatcattct 720

ttcttattgc ttatctctaa agcaaaaact tccctaaacc ctaagccaaa ggactcagat 780

cgatgcagaa ccaagaaggc ttgttttgga tttgagagcc aaatgcaaag aaaaaaactc 840

tt 842

<210> 19
<211> 1130
<212> PRT
<213> Arabidopsis thaliana

<400> 19

Met Pro Ile Pro Arg Lys Lys Ala Leu Thr Val Ser His Phe Ser Ile

1 5 10 15

Thr Leu Ser Leu Phe Leu Ala Phe Phe Ile Ser Ser Thr Ser Ala Ser
20 25 30

Thr Asn Glu Val Ser Ala Leu Ile Ser Trp Leu His Ser Ser Asn Ser
35 40 45

Pro Pro Pro Ser Val Phe Ser Gly Trp Asn Pro Ser Asp Ser Asp Pro
50 55 60

Cys Gln Trp Pro Tyr Ile Thr Cys Ser Ser Pro Asp Asn Lys Leu Val
65 70 75 80

Thr Glu Ile Asn Val Val Ser Val Gln Leu Ala Leu Pro Phe Pro Pro
85 90 95

Asn Ile Ser Ser Phe Thr Ser Leu Gln Lys Leu Val Ile Ser Asn Thr
100 105 110

Asn Leu Thr Gly Ala Ile Ser Ser Glu Ile Gly Asp Cys Ser Glu Leu
115 120 125

Ile Val Ile Asp Leu Ser Ser Asn Ser Leu Val Gly Glu Ile Pro Ser
130 135 140

Ser Leu Gly Lys Leu Lys Asn Leu Gln Glu Leu Cys Leu Asn Ser Asn
145 150 155 160

Gly Leu Thr Gly Lys Ile Pro Pro Glu Leu Gly Asp Cys Val Ser Leu
165 170 175

Lys Asn Leu Glu Ile Phe Asp Asn Tyr Leu Ser Glu Asn Leu Pro Leu
180 185 190

Glu Leu Gly Lys Ile Ser Thr Leu Glu Ser Ile Arg Ala Gly Gly Asn
195 200 205

Ser Glu Leu Ser Gly Lys Ile Pro Glu Glu Ile Gly Asn Cys Arg Asn
210 215 220

Leu Lys Val Leu Gly Leu Ala Ala Thr Lys Ile Ser Gly Ser Leu Pro
225 230 235 240

Val Ser Leu Gly Gln Leu Ser Lys Leu Gln Ser Leu Phe Val Tyr Ser
245 250 255

Thr Ser Leu Ser Gly Glu Ile Pro Lys Glu Leu Gly Asn Cys Ser Glu
260 265 270

Leu Ile Asn Leu Phe Leu Tyr Asp Asn Asp Leu Ser Gly Thr Leu Pro
275 280 285

Lys Glu Leu Gly Lys Leu Gln Asn Leu Glu Lys Met Leu Leu Trp Gln
290 295 300

Asn Asn Leu His Gly Pro Ile Pro Glu Glu Ile Gly Phe Met Lys Ser
305 310 315 320

Leu Asn Ala Ile Asp Leu Ser Met Asn Tyr Phe Ser Gly Thr Ile Pro
325 330 335

Lys Ser Phe Gly Asn Leu Ser Asn Leu Gln Glu Leu Met Leu Ser Ser
340 345 350

Asn Asn Ile Thr Gly Ser Ile Pro Ser Ile Leu Ser Asp Cys Thr Lys
355 360 365

Leu Val Gln Phe Gln Ile Asp Ala Asn Gln Ile Ser Gly Leu Ile Pro
370 375 380

Pro Thr Ile Gly Leu Leu Lys Glu Leu Asn Ile Phe Gly Trp Gln Asn
385 390 395 400

Lys Leu Glu Gly Asn Ile Pro Asp Glu Leu Ala Gly Cys Gln Asn Leu
405 410 415

Gln Ala Leu Asp Leu Ser Gln Asn Tyr Leu Thr Gly Ser Leu Pro Ala
420 425 430

Gly Leu Phe Gln Leu Arg Asn Leu Thr Lys Leu Leu Leu Ile Ser Asn
435 440 445

Ala Ile Ser Gly Val Ile Pro Leu Glu Thr Gly Asn Cys Thr Ser Leu
450 455 460

Val Arg Leu Arg Leu Val Asn Asn Arg Ile Thr Gly Glu Ile Pro Lys
465 470 475 480

Gly Ile Leu Phe Leu Gln Asn Leu Ser Phe Leu Asp Leu Ser Glu Asn
485 490 495

Asn Leu Ser Gly Pro Val Pro Leu Glu Ile Ser Asn Cys Arg Gln Leu
500 505 510

Gln Met Leu Asn Leu Ser Asn Asn Thr Leu Gln Gly Tyr Leu Pro Leu
515 520 525

Ser Leu Ser Ser Leu Thr Lys Leu Gln Val Leu Asp Val Ser Ser Asn
530 535 540

Asp Leu Thr Gly Lys Ile Pro Asp Ser Leu Gly His Leu Ile Ser Leu
545 550 555 560

Asn Arg Leu Ile Leu Ser Lys Asn Ser Phe Asn Gly Glu Ile Pro Ser
565 570 575

Ser Leu Gly His Cys Thr Asn Leu Gln Leu Leu Asp Leu Ser Ser Asn
580 585 590

Asn Ile Ser Gly Thr Ile Pro Glu Glu Leu Phe Asp Ile Gln Asp Leu
595 600 605

Asp Ile Ala Leu Asn Leu Ser Trp Asn Ser Leu Asp Gly Phe Ile Pro
610 615 620

Glu Arg Ile Ser Ala Leu Asn Arg Leu Ser Val Leu Asp Ile Ser His
625 630 635 640

Asn Met Leu Ser Gly Asp Leu Ser Ala Leu Ser Gly Leu Glu Asn Leu
645 650 655

Val Ser Leu Asn Ile Ser His Asn Arg Phe Ser Gly Tyr Leu Pro Asp
660 665 670

Ser Lys Val Phe Arg Gln Leu Ile Gly Ala Glu Met Glu Gly Asn Asn
675 680 685

Gly Leu Cys Ser Lys Gly Phe Arg Ser Cys Phe Val Ser Asn Ser Ser
690 695 700

Gln Leu Thr Thr Gln Arg Gly Val His Ser His Arg Leu Arg Ile Ala
705 710 715 720

Ile Gly Leu Leu Ile Ser Val Thr Ala Val Leu Ala Val Leu Gly Val
725 730 735

Leu Ala Val Ile Arg Ala Lys Gln Met Ile Arg Asp Asp Asn Asp Ser
740 745 750

Glu Thr Gly Glu Asn Leu Trp Thr Trp Gln Phe Thr Pro Phe Gln Lys
755 760 765

Leu Asn Phe Thr Val Glu His Val Leu Lys Cys Leu Val Glu Gly Asn
770 775 780

Val Ile Gly Lys Gly Cys Ser Gly Ile Val Tyr Lys Ala Glu Met Pro
785 790 795 800

Asn Arg Glu Val Ile Ala Val Lys Lys Leu Trp Pro Val Thr Val Pro
805 810 815

Asn Leu Asn Glu Lys Thr Lys Ser Ser Gly Val Arg Asp Ser Phe Ser
820 825 830

Ala Glu Val Lys Thr Leu Gly Ser Ile Arg His Lys Asn Ile Val Arg
835 840 845

Phe Leu Gly Cys Cys Trp Asn Lys Asn Thr Arg Leu Leu Met Tyr Asp
850 855 860

Tyr Met Ser Asn Gly Ser Leu Gly Ser Leu Leu His Glu Arg Ser Gly
865 870 875 880

Val Cys Ser Leu Gly Trp Glu Val Arg Tyr Lys Ile Ile Leu Gly Ala
885 890 895

Ala Gln Gly Leu Ala Tyr Leu His His Asp Cys Val Pro Pro Ile Val
900 905 910

His Arg Asp Ile Lys Ala Asn Asn Ile Leu Ile Gly Pro Asp Phe Glu
915 920 925

Pro Tyr Ile Gly Asp Phe Gly Leu Ala Lys Leu Val Asp Asp Gly Asp
930 935 940

Phe Ala Arg Ser Ser Asn Thr Ile Ala Gly Ser Tyr Gly Tyr Ile Ala
945 950 955 960

Pro Glu Tyr Gly Tyr Ser Met Lys Ile Thr Glu Lys Ser Asp Val Tyr
965 970 975

Ser Tyr Gly Val Val Val Leu Glu Val Leu Thr Gly Lys Gln Pro Ile
980 985 990

Asp Pro Thr Ile Pro Asp Gly Leu His Ile Val Asp Trp Val Lys Lys
995 1000 1005

Ile Arg Asp Ile Gln Val Ile Asp Gln Gly Leu Gln Ala Arg Pro

Ile Glu 1015 1020

Glu Ser Glu Val Glu Glu Met Met Gln Thr Leu Gly Val Ala Leu
1030 1035

Leu Cys Ile Asn Pro Ile Pro Glu Asp Arg Pro Thr Met Lys Asp
1045 1050

Val Ala Ala Met Leu Ser Glu Ile Cys Gln Glu Arg Glu Glu Ser
1060 1065

Met Lys Val Asp Gly Cys Ser Gly Ser Cys Asn Asn Gly Arg Glu
1075 1080

Arg Gly Lys Asp Asp Ser Thr Ser Ser Val Met Gln Gln Thr Ala
1090 1095

Lys Tyr Leu Arg Ser Ser Ser Thr Ser Phe Ser Ala Ser Ser Leu
1105 1110

Leu Ser Ser Ser Ser Ser Ser Ala Thr Ser Asn Val Arg Pro Asn
1120 1125

Leu Lys
1140